

# **ANALISIS SENTIMEN OPINI PELANGGAN TERHADAP ASPEK PARIWISATA PANTAI MALANG SELATAN MENGGUNAKAN TF-IDF DAN SUPPORT VECTOR MACHINE**

## **SKRIPSI**

Untuk memenuhi sebagian persyaratan  
memperoleh gelar Sarjana Komputer

Disusun oleh:  
Yoga Tika Pratama  
NIM: 145150400111003



PROGRAM STUDI SISTEM INFORMASI  
JURUSAN SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS BRAWIJAYA  
MALANG  
2018

## PENGESAHAN

ANALISIS SENTIMEN OPINI PELANGGAN TERHADAP ASPEK PARIWISATA PANTAI  
MALANG SELATAN MENGGUNAKAN TF-IDF DAN SUPPORT VECTOR MACHINE

### SKRIPSI

Diajukan untuk memenuhi sebagian persyaratan  
memperoleh gelar Sarjana Komputer

Disusun Oleh :  
Yoga Tika Pratama  
NIM: 145150400111003

Skrripsi ini telah diuji dan dinyatakan lulus pada  
1 Agustus 2018  
Telah diperiksa dan disetujui oleh:

Dosen Pembimbing I

*Fitra A Bachtiar*

Dr. Eng. Fitra A. Bachtiar, S.T., M.Eng.  
NIK. 201201 840628 1 001

Dosen Pembimbing II

*Nanang Yudi Setiawan*

Nanang Yudi Setiawan, S.T., M.Kom.  
NIP. 19760619 200604 1 001

Mengetahui

Ketua Jurusan Sistem Informasi



*Herman Tolle*  
Dr. Eng. Herman Tolle, S.T., M.T.  
NIP. 19740823 200012 1 001

## PERNYATAAN ORISINALITAS

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah skripsi ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini dan disebutkan dalam daftar pustaka.

Apabila ternyata didalam naskah skripsi ini dapat dibuktikan terdapat unsur-unsur plagiasi, saya bersedia skripsi ini digugurkan dan gelar akademik yang telah saya peroleh (sarjana) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).

Malang, 20 Juli 2018

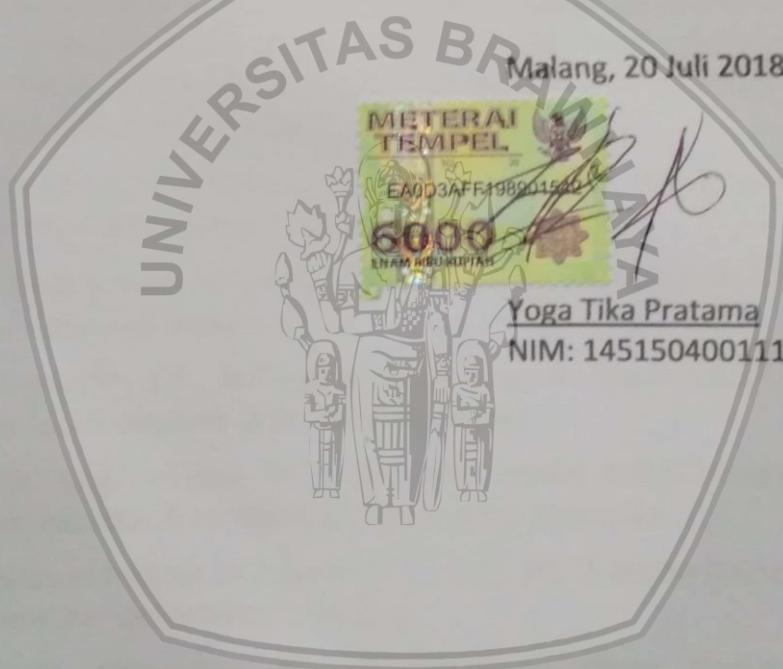
METERAI  
TEMPEL

EA0D3AFF498001502

6000  
ENAM RIBU RUPIAH

Yoga Tika Pratama

NIM: 145150400111003



## KATA PENGANTAR

Syukur Alhamdulillah dihaturkan kepada Allah Subhanahu wa Ta'ala atas segala berkah, rahmat, dan hidayah yang senantiasa diberikan sehingga penulis dapat menyelesaikan skripsi ini yang berjudul “Analisis Sentimen Opini Pelanggan Terhadap Aspek Pariwisata Pantai Malang Selatan Menggunakan TF-IDF Dan *Support Vector Machine*”. Shalawat dan salam semoga dicurahkan kepada junjungan dan suri tauladan kita, Nabi Muhammad Shalallahu 'Alaihi Wassalam yang telah memberikan tuntunan dan petunjuk kepada umat manusia.

Penulisan skripsi ini merupakan salah satu syarat bagi penulis dalam memperoleh gelar Sarjana Komputer pada Fakultas Ilmu Komputer Universitas Brawijaya. Dalam penulisan skripsi ini penulis banyak belajar berbagai hal yang tidak diajarkan pada materi perkuliahan. Belajar menghadapi permasalahan dan mengatasinya serta rasa semangat yang penulis miliki sehingga dapat menyelesaikan skripsi ini. Pada kesempatan kali ini, penulis ingin menyampaikan rasa syukur dan terima kasih kepada :

1. Bapak Dr. Eng. Fitra Abdurrachman Bachtiar, S.T, M.Eng selaku dosen Pembimbing I dalam penyusunan skripsi.
2. Bapak Nanang Yudi Setiawan, S.T., M.Kom selaku dosen Pembimbing II dalam penyusunan skripsi.
3. Bapak Suprpto, S.T., M.T selaku ketua Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya.
4. Bapak Dr. Eng. Herman Tolle, S.T., M.T, selaku Ketua Jurusan Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya.
5. Bapak Wayan Firdaus Mahmudy, S.Si., M.T., Ph. D selaku Dekan Fakultas Ilmu Komputer Universitas Brawijaya.
6. Pemerintah Kabupaten Malang khususnya Dinas Pariwisata dan Kebudayaan Kabupaten Malang yang telah bersedia bekerja sama serta membantu penulis dalam proses penelitian.
7. Keluarga penulis, Bapak Sartika dan Ibu Toenah sebagai orang tua yang selalu memberikan dukungan moril dan materiil serta motivasi, nasihat, dan doa yang tulus. Adik Surya Dermawan yang berkenan menjaga kedua orang tua selagi penulis melaksanakan kuliah.
8. Teman seperjuangan, Dwi Asri Nuryulianti yang selalu memberikan semangat, kritik, saran, dan waktu berharganya kepada penulis dalam melakukan pengerjaan skripsi.
9. Keluarga Tim Bujang. Iqbal, Wahyu, Vicky, Ferlie, dan Wendi yang berkenan menjadi tempat untuk saling mengembangkan diri dan berbagi mimpi bersama.

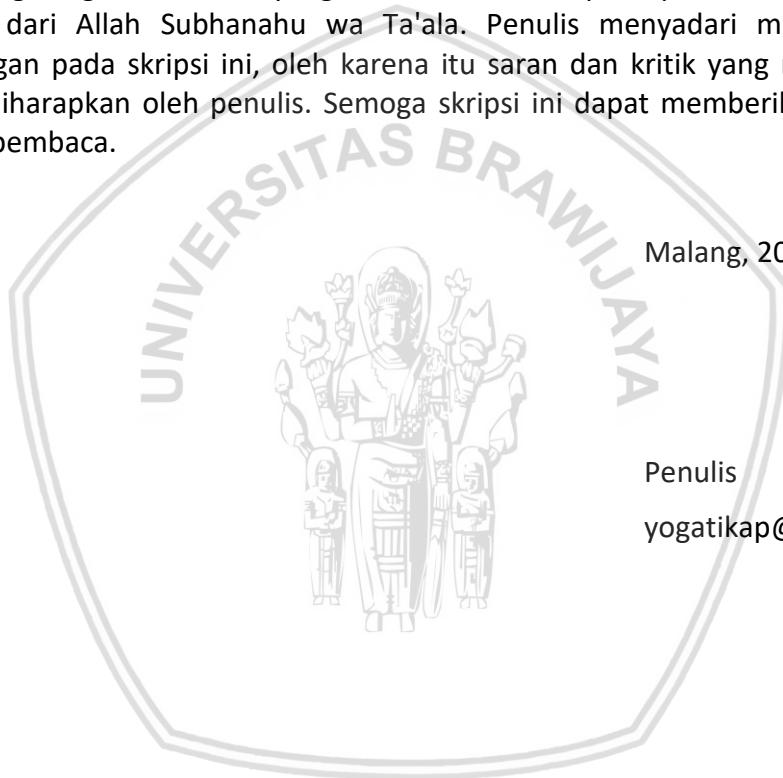


10. Teman-teman Distrik Gading 33. Mas Fadhil, Mas Galih, Mas Yoto, Mas Johan, dan Mas Rama.
11. Teman-teman penulis dari Gerakan 10 September dan Bumblebee. Terima kasih atas segala pengalaman berharga yang kalian berikan.
12. Segenap civitas Fakultas Ilmu Komputer Universitas Brawijaya, khususnya teman-teman Jurusan Sistem Informasi Angkatan 2014. Terimakasih atas pengalaman berharga dan bantuan yang telah diberikan kepada penulis semasa perkuliahan.
13. Semua pihak yang tidak bisa penulis sebutkan satu persatu yang telah membantu penulis dalam menyelesaikan skripsi ini.

Semoga segala kebaikan yang telah diberikan kepada penulis mendapatkan balasan dari Allah Subhanahu wa Ta'ala. Penulis menyadari masih banyak kekurangan pada skripsi ini, oleh karena itu saran dan kritik yang membangun sangat diharapkan oleh penulis. Semoga skripsi ini dapat memberikan manfaat kepada pembaca.

Malang, 20 Juli 2018

Penulis  
yogatikap@gmail.com



## ABSTRAK

Dinas Pariwisata dan Kebudayaan (Disparbud) Kabupaten Malang membutuhkan perspektif pelanggan dalam proses pengawasan dan evaluasi kegiatan pengelolaan dan pengembangan pariwisata pantai Malang Selatan. Akan tetapi Disparbud tidak memiliki data tentang perspektif pelanggan, maupun penerapan teknologi pengolahan dan analisis data yang mampu memberikan informasi tentang perspektif pelanggan terhadap aspek pariwisata pantai Malang Selatan. Perspektif pelanggan bisa didapatkan dari opini pelanggan dalam bentuk ulasan yang ada pada *platform Web*. Metode *Web Scraping* adalah salah satu metode untuk melakukan ekstraksi data opini pelanggan yang bersumber dari *platform Web*. Untuk mendapatkan perspektif pelanggan terhadap aspek pariwisata pantai Malang Selatan bisa dengan melakukan Analisis Sentimen pada tingkat aspek. Salah satu metode klasifikasi yaitu *Support Vector Machine* yang dapat digunakan untuk melakukan klasifikasi sentimen dalam proses Analisis Sentimen tersebut. Pada Penelitian ini berhasil mendapatkan 674 data opini pelanggan dalam Bahasa Indonesia dari tahun 2013 hingga tahun 2018 untuk 43 objek pariwisata pantai Malang Selatan yang bersumber dari situs TripAdvisor. Pengujian terhadap hasil klasifikasi sentimen terhadap setiap aspek menunjukkan hasil yang baik dengan rata-rata *Accuracy* sebesar 85%, *Precision* sebesar 85%, *Recall* sebesar 87%, dan *F1-Score* sebesar 85%. Luaran dari penelitian ini berupa visualisasi data dalam bentuk *Dashboard* yang terdiri dari 4 visualisasi *Dashboard* yang memuat hasil dari Analisis Sentimen opini pelanggan terhadap aspek pariwisata pantai Malang Selatan.

**Kata kunci:** *Web Scraping*, Analisis Sentimen, Klasifikasi, SVM, Visualisasi Data

## ABSTRACT

Department of Tourism and Culture (Disparbud) of Malang Regency requires the customer's perspective in the process of monitoring and evaluation of management and development activities of South Malang beach tourism. But Disparbud neither have the data about the customer's opinion, nor the application of technology on data processing and data analysis that capable of providing information about customer's perspective on aspects of South Malang beach tourism. The customer's perspective can be obtained from customer's opinions in the form of review that available on the Web platform. Web Scraping is a method for extraction of customer's opinion data sourced from the Web platform. To get the information about customer's perspective on aspects of South Malang beach tourism could be done by performing Aspect Level Sentiment Analysis on the customer's opinion data. One of many classification methods is the Support Vector Machine, which can be used to classify sentiments in Sentiment Analysis process. This research managed to get 674 customer's opinion data in Bahasa Indonesia ranged from 2013 to 2018 for 43 tourist attractions of South Malang beach tourism sourced from TripAdvisor. The tests that have been conducted on the results of the sentiment classification for all aspects resulted good results in an average of 85% on Accuracy, 85% on Precision, 87% on Recall, and 85% on F1-Score. The output of this research is in form of data visualization in the form of Dashboard consisted of 4 Dashboards which contains results from the process of Sentiment Analysis on the customer opinions to aspects of South Malang beach tourism.

**Keywords:** Web Scraping, Sentiment Analysis, Classification, SVM, Data Visualization

## DAFTAR ISI

PENGESAHAN .....	ii
PERNYATAAN ORISINALITAS .....	iii
KATA PENGANTAR.....	iv
ABSTRAK.....	vi
ABSTRACT .....	vii
DAFTAR ISI .....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR SOURCE CODE .....	xiii
DAFTAR LAMPIRAN .....	xiv
<b>BAB 1 PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar belakang.....	1
1.2 Rumusan masalah.....	3
1.3 Tujuan .....	4
1.4 Manfaat.....	4
1.5 Batasan masalah .....	4
1.6 Sistematika pembahasan .....	5
<b>BAB 2 LANDASAN KEPUSTAKAAN .....</b>	<b>6</b>
2.1 Kajian pustaka .....	6
2.2 Profil Dinas Pariwisata dan Kebudayaan Kabupaten Malang.....	8
2.2.1 Pengawasan dan evaluasi kegiatan pengelolaan pariwisata .....	8
2.3 Opini.....	9
2.3.1 Teori Opini.....	9
2.3.2 Komponen Opini .....	9
2.4 <i>Web Scraping</i> .....	10
2.4.1 WebHarvy.....	10
2.5 <i>Data Mining</i> .....	11
2.5.1 Definisi <i>Data Mining</i> .....	11
2.5.2 Tahapan dalam <i>Data Mining</i> .....	11
2.5.3 Teknik analisis data dalam <i>Data Mining</i> .....	12
2.5.4 <i>Text Mining</i> .....	13
2.6 Analisis Sentimen .....	14
2.6.2 Kerangka kerja Analisis Sentimen .....	15
2.7 <i>Term Weighting</i> .....	16
2.7.1 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> .....	16
2.8 <i>Machine Learning</i> .....	17
2.8.1 <i>Support Vector Machine (SVM)</i> .....	17
2.9 Pemilihan dan evaluasi model .....	19
2.9.1 <i>Kfold Cross-Validation</i> .....	19
2.9.2 Metriks untuk evaluasi model.....	20
2.10 Bahasa Pemrograman Python .....	21
2.10.1 <i>Nature Language ToolKit (NLTK)</i> .....	21
2.10.2 <i>Sastrawi</i> .....	22



2.10.3 Scikit-learn .....	22
2.11 Visualisasi .....	22
2.11.1 Dashboard .....	22
BAB 3 METODOLOGI .....	24
3.1 Metodologi penelitian .....	24
3.1.1 Identifikasi masalah .....	24
3.1.2 Studi literatur .....	24
3.1.3 Pengumpulan data .....	24
3.1.4 Text Preprocessing .....	25
3.1.5 Term Weighting.....	25
3.1.6 Klasifikasi sentimen .....	25
3.1.7 Analisis dan pengujian hasil klasifikasi sentimen .....	26
3.1.8 Pembuatan visualisasi hasil Analisis Sentimen .....	26
3.1.9 Penarikan kesimpulan dan saran .....	26
BAB 4 PERANCANGAN.....	27
4.1 Pengumpulan data .....	27
4.2 Pemilihan aspek pariwisata pantai Malang Selatan .....	34
4.3 Perancangan <i>dataset</i> .....	34
4.4 Desain antar muka <i>Dashboard</i> .....	39
BAB 5 IMPLEMENTASI .....	43
5.1 Text Preprocessing .....	43
5.1.1 Pembuatan indeks kata .....	43
5.1.2 Case folding .....	44
5.1.3 Cleansing .....	45
5.1.4 Tokenisasi .....	46
5.1.5 Formalisasi dan translasi .....	47
5.1.6 Stemming .....	47
5.1.7 Stop-word Removal .....	49
5.2 Term Weighting .....	50
5.2.1 Manualisasi .....	50
5.3 Klasifikasi sentimen.....	51
5.3.1 Pemilihan model menggunakan <i>K-fold Cross Validation</i> .....	52
5.3.2 Klasifikasi sentimen menggunakan <i>LinearSVC</i> .....	53
BAB 6 ANALISIS DAN PEMBAHASAN .....	54
6.1 Hasil klasifikasi .....	54
6.1.1 Aspek Umum .....	54
6.1.2 Aspek Kebersihan .....	57
6.1.3 Aspek Keramaian.....	60
6.1.4 Aspek Akses Jalan.....	63
6.1.5 Aspek Ombak .....	66
6.2 Pengujian hasil klasifikasi sentimen.....	69
6.3 Visualisasi <i>Dashboard</i> .....	70
6.3.1 <i>Dashboard</i> Analisis Sentimen .....	70
6.3.2 <i>Dashboard trend</i> sentimen .....	70
6.3.3 <i>Dashboard</i> objek wisata.....	70

6.3.4 <i>Dashboard</i> aspek.....	71
6.4 Analisis hasil <i>Usability Testing Dashboard</i> .....	75
BAB 7 PENUTUP .....	77
7.1 Kesimpulan.....	77
7.2 Saran .....	78
DAFTAR PUSTAKA.....	79
LAMPIRAN A Contoh Data Mentah hasil <i>Scraping</i> .....	83
LAMPIRAN B Contoh Hasil pelabelan Data .....	85
LAMPIRAN C Tampilan <i>Dashboard</i> Aspek.....	87
LAMPIRAN D Transkrip Wawancara Dan Kuesioner SUS.....	92



## DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i> .....	20
Tabel 4.1 Variabel dan tipe data hasil proses <i>Scraping</i> .....	32
Tabel 4.2 Contoh data hasil proses <i>Scraping</i> .....	33
Tabel 4.3 Deskripsi aspek pariwisata pantai Malang Selatan .....	34
Tabel 4.4 Polaritas sentimen aspek pariwisata pantai Malang Selatan .....	34
Tabel 4.5 Pelabelan data untuk aspek Umum .....	35
Tabel 4.6 Pelabelan data untuk aspek Kebersihan .....	36
Tabel 4.7 Pelabelan data untuk aspek Keramaian .....	36
Tabel 4.8 Pelabelan data untuk aspek Akses Jalan .....	37
Tabel 4.9 Pelabelan data untuk aspek akses Ombak .....	38
Tabel 5.1 Hasil implementasi <i>Case folding</i> .....	44
Tabel 5.2 Hasil implementasi <i>Cleansing</i> .....	45
Tabel 5.3 Hasil implementasi tokenisasi .....	46
Tabel 5.4 Formalisasi kata .....	47
Tabel 5.5 Translasi kata .....	47
Tabel 5.6 Implementasi <i>Stemming</i> .....	48
Tabel 5.7 Implementasi <i>Stop-word removal</i> .....	49
Tabel 5.8 Menghitung TF, DF, dan IDF .....	50
Tabel 5.9 Menghitung bobot TF-IDF .....	51
Tabel 6.1 Hasil klasifikasi aspek Umum .....	54
Tabel 6.2 Hasil klasifikasi aspek Kebersihan .....	57
Tabel 6.3 Hasil klasifikasi aspek Keramaian .....	60
Tabel 6.4 Hasil klasifikasi aspek Akses Jalan .....	63
Tabel 6.5 Hasil klasifikasi aspek Ombak .....	66
Tabel 6.6 Hasil pengujian akurasi .....	69
Tabel 6.7 Hasil penilaian <i>System Usability Scale</i> .....	75

## DAFTAR GAMBAR

Gambar 2.1 <i>Optimal Hyperplane</i> .....	17
Gambar 2.2 <i>Support Vector</i> .....	19
Gambar 2.3 Ilustrasi <i>Stratified 5 fold Cross Validation</i> .....	20
Gambar 3.1 Metodologi Penelitian .....	24
Gambar 4.1 Ulasan pada situs TripAdvisor .....	27
Gambar 4.2 <i>Load</i> target URL pada WebHarvy .....	28
Gambar 4.3 Konfigurasi <i>next page</i> .....	28
Gambar 4.4 Konfigurasi <i>Scraping</i> variabel Judul .....	29
Gambar 4.5 Konfigurasi <i>Scraping</i> variabel Tanggal .....	29
Gambar 4.6 Konfigurasi <i>Scraping</i> variabel Nama .....	30
Gambar 4.7 Konfigurasi <i>Scraping</i> variabel Teks .....	30
Gambar 4.8 <i>Preview</i> data hasil konfigurasi <i>Scraping</i> .....	31
Gambar 4.9 Antar muka jendela proses <i>Scraping</i> pada WebHarvy .....	31
Gambar 4.10 Tampilan jendela proses <i>Scraping</i> setelah proses selesai .....	32
Gambar 4.11 Penyimpanan data hasil <i>Scraping</i> .....	32
Gambar 4.12 Desain Dashboard Analisis Sentimen .....	41
Gambar 4.13 Desain antarmuka <i>Dashboard</i> tren perubahan sentimen .....	41
Gambar 4.14 Desain antarmuka <i>Dashboard</i> pemeringkatan objek wisata .....	42
Gambar 4.15 Desain antarmuka <i>Dashboard</i> pemeringkatan objek wisata .....	42
Gambar 5.1 Tahapan <i>Text preprocessing</i> .....	43
Gambar 6.1 Frekuensi hasil klasifikasi terhadap aspek Umum .....	55
Gambar 6.2 Tren hasil klasifikasi terhadap aspek Umum .....	56
Gambar 6.3 Ulasan berdasarkan objek wisata terhadap aspek Umum .....	56
Gambar 6.4 Frekuensi hasil klasifikasi terhadap aspek Kebersihan .....	58
Gambar 6.5 Tren hasil klasifikasi terhadap aspek Kebersihan .....	59
Gambar 6.6 Ulasan berdasarkan objek wisata terhadap aspek Kebersihan .....	59
Gambar 6.7 Frekuensi hasil klasifikasi terhadap aspek Keramaian .....	61
Gambar 6.8 Tren hasil klasifikasi terhadap aspek Keramaian .....	62
Gambar 6.9 Ulasan berdasarkan objek wisata terhadap aspek Keramaian .....	62
Gambar 6.10 Frekuensi hasil klasifikasi terhadap aspek Akses Jalan .....	64
Gambar 6.11 Tren hasil klasifikasi terhadap aspek Akses Jalan .....	65
Gambar 6.12 Ulasan berdasarkan objek wisata terhadap aspek Akses Jalan .....	65
Gambar 6.13 Frekuensi hasil klasifikasi terhadap aspek Ombak .....	67
Gambar 6.14 Tren hasil klasifikasi terhadap aspek Ombak .....	68
Gambar 6.15 Ulasan berdasarkan objek wisata terhadap aspek Ombak .....	68
Gambar 6.16 Dashboard Analisis Sentimen .....	72
Gambar 6.17 <i>Dashboard Trend</i> Sentimen .....	73
Gambar 6.18 <i>Dashboard</i> Objek Wisata .....	74
Gambar 6.19 Skala pemeringkatan <i>SUS Score</i> .....	76



## DAFTAR SOURCE CODE

<i>Source code 5.1 Implementasi pembuatan indeks kata (Term indexing)</i> .....	44
<i>Source code 5.2 Implementasi Case folding</i> .....	44
<i>Source code 5.3 Implementasi Cleansing</i> .....	45
<i>Source code 5.4 Implementasi tokenisasi</i> .....	46
<i>Source code 5.5 Implementasi Stemming</i> .....	48
<i>Source code 5.6 Implementasi Stop-word Removal</i> .....	49
<i>Source code 5.7 Implementasi Term Weighting menggunakan TF-IDF</i> .....	50
<i>Source code 5.8 Implementasi Stratified K-Fold Cross Validation</i> .....	53
<i>Source code 5.9 Implementasi klasifikasi sentimen menggunakan LinearSVC</i> .....	53
<i>Source code 6.1 Source code pengujian klasifikasi sentimen</i> .....	69



## DAFTAR LAMPIRAN

LAMPIRAN A Contoh data mentah hasil <i>Scraping</i> .....	83
LAMPIRAN B Contoh hasil pelabelan data .....	85
LAMPIRAN C Tampilan <i>Dashboard</i> aspek .....	87
LAMPIRAN D Transkrip wawancara dan kuesioner SUS .....	92



# CURRICULUM VITAE

## **PERSONAL DETAILS**

Name : Yoga Tika Pratama  
Place, DOB : Indramayu, 28 Mei 1996  
Gender : Male  
Nationality : Indonesia  
Current Address : Gading Street 33 Klojen- Malang  
Phone (mobile) : +62895635306941  
E-mail : [yogatikap@gmail.com](mailto:yogatikap@gmail.com)



## **FORMAL EDUCATION**

2014- Brawijaya University  
2011 – 2014 SMKN 2 Kota Bekasi  
2008 – 2011 SMPN 1 Cileungsi  
2002 – 2008 SDN 2 Cinyosog

## **ACHIEVEMENTS**

Finalist. 2016 Brawijaya Young Entrepreneur Olympiad

## **WORKING EXPERIENCE**

2013 Internship Junior IT Network Support, PT Nandya Karya Perkasa. Bogor, West Java, Indonesia.  
2017 Internship Student Lentera Bumi Nusantara, Tasikmalaya, West Jawa, Indonesia.

## BAB 1 PENDAHULUAN

### 1.1 Latar belakang

Pada saat ini sektor pariwisata merupakan salah satu sektor industri yang menjanjikan. Hal tersebut terbukti dengan nilai perputaran uang yang mencapai hingga 1,235 Milyar Dollar Amerika pada tahun 2016 (World Tourism Organization, 2017). Bahkan World Tourism Organization juga memperkirakan bahwa angka kunjungan wisatawan dunia akan naik sebesar 200% pada tahun 2020 (Utama, 2016). Pemerintah Republik Indonesia melalui Kementerian Pariwisata, memasang target tinggi untuk sektor pariwisata. Diharapkan pada tahun 2019, Indonesia bisa menerima 240 Trilyun rupiah dari kunjungan wisatawan mancanegara (Kementerian Pariwisata Republik Indonesia, 2014). Kondisi ini mengakibatkan pemerintah pada tingkat daerah melakukan pembenahan dalam bidang pariwisata di daerahnya masing-masing demi tercapainya citra pariwisata yang baik serta meningkatkan kunjungan wisatawan.

Berkembangnya *Web 2.0* yang terjadi saat ini berorientasi pada *User Generated Content*. Media sosial seperti *blog*, forum, atau platform jejaring sosial seperti Facebook dan Twitter telah menjadi bagian penting dalam hidup manusia, media sosial menjadi tempat di mana seorang individu dapat berbagi opini dan informasi (Pozzi et al., 2016). Hal ini turut juga dimanfaatkan oleh para wisatawan dengan menggunakan media sosial sebagai wadah untuk mengutarakan ulasan terkait perjalanan wisata yang telah dilaluinya. Selain media sosial yang disebutkan sebelumnya, ada pula situs *Web* seperti salah satu contohnya ialah TripAdvisor yang menyediakan fitur khusus di mana pengguna dapat memberikan ulasan dan penilaian terkait suatu objek pariwisata.

Fenomena banyaknya ulasan yang diberikan pada platform *Web* menghasilkan volume data yang besar terkait dengan opini terhadap suatu objek pariwisata. Data tersebut perlu untuk diakuisisi untuk kemudian diolah dan dianalisis untuk dapat menghasilkan informasi yang berguna bagi pemegang keputusan. Salah satu cara untuk menganalisis yakni dengan menerapkan Analisis Sentimen. Analisis sentimen menggunakan *raw data* yang bersumber dari media sosial dan menggali pengetahuan dari data tersebut untuk kemudian digunakan sebagai dasar dalam pengambilan keputusan strategis (Wright, 2009).

Pemerintah Daerah (Pemda) Kabupaten Malang sangat memperhatikan sektor pariwisata. Hal tersebut tercermin dengan adanya *branding* "*Kabupaten Malang: The Heart of East Java*", yang di dalamnya menyebutkan bahwa optimalisasi pariwisata menjadi salah satu dari tiga prioritas utama pembangunan Kabupaten Malang. Kabupaten Malang memiliki Dinas khusus untuk menangani sektor pariwisata, dinas tersebut adalah Dinas Pariwisata dan Kebudayaan Kabupaten Malang (Disparbud). Kabupaten Malang memiliki 169 objek pariwisata yang terbagi ke dalam 3 kategori yakni objek pariwisata alam, objek pariwisata



budaya, dan objek pariwisata buatan. Pariwisata yang menjadi andalan Kabupaten Malang adalah wisata pantai dengan jumlah 44 destinasi wisata pantai yang berjajar di sepanjang daerah Malang Selatan. Pelanggan pariwisata Kabupaten Malang di dominasi oleh wisatawan lokal dengan 5,7 juta kunjungan dari total 5,9 Juta kunjungan wisatawan ke Kabupaten Malang pada tahun 2016 (Dinas Pariwisata dan Kebudayaan Kabupaten Malang, 2018).

Disparbud Kabupaten Malang melalui Perusahaan Daerah (PD) Jasa Yasa dalam menjalankan fungsinya memiliki strategi yang dijalankan dalam 3 perspektif yakni perspektif sumber daya, pelanggan, dan keuangan. Perspektif sumber daya yang dimiliki, yaitu meliputi pemeliharaan, penataan dan penciptaan sarana penunjang, memberikan rasa aman dan nyaman juga melaksanakan pengawasan dan meningkatkan sumber daya manusia. Perspektif pelanggan bertujuan untuk memberikan kepuasan dan kenyamanan dengan tujuan meningkatkan dan memaksimalkan pemenuhan kebutuhan pelanggan. Perspektif keuangan, dengan tujuan terciptanya kondisi keuangan yang kuat sehingga nilai perusahaan semakin baik dan memberikan tingkat kepuasan yang tinggi bagi stakeholder. (PD Jasa Yasa, 2013). Disparbud Kabupaten Malang selaku pembina PD Jasa Yasa, bertanggung jawab dalam melakukan pengawasan dan evaluasi terhadap kegiatan manajemen dan pengembangan pariwisata yang berorientasi pada pemenuhan ketiga perspektif tersebut.

Berdasarkan hasil wawancara dengan Disparbud Kabupaten Malang didapatkan beberapa permasalahan dalam hal penerapan strategi yang bertujuan untuk memenuhi perspektif pelanggan. Permasalahan tersebut antara lain Disparbud memiliki kesulitan dalam upaya pengembangan pariwisata pantai Malang Selatan yang dapat memenuhi tujuan strategi dalam perspektif pelanggan. Selain itu, Disparbud juga tidak memiliki data mengenai perspektif pelanggan maupun penerapan teknologi pengolahan dan analisis data yang mampu memberikan informasi tentang perspektif pelanggan terhadap pariwisata pantai Malang Selatan. Selain itu, penerapan teknologi komputer sebatas pada menyimpan data, sehingga belum terdapat visualisasi yang mampu menggambarkan perspektif pelanggan terhadap aspek pariwisata pantai Malang Selatan guna mempermudah Disparbud dalam melakukan upaya pengembangan pariwisata pantai Malang Selatan yang dapat memenuhi tujuan strategi dalam perspektif pelanggan.

Dari masalah yang dijelaskan pada Disparbud Kabupaten Malang maka perlu untuk mendapatkan data opini pelanggan berbahasa Indonesia. Kemudian melakukan Analisis Sentimen terhadap data yang berisi opini pelanggan tersebut untuk dapat mengetahui perspektif pelanggan terhadap aspek objek pariwisata pantai Malang Selatan. Sehingga hasil pengolahan dan analisis data kemudian dapat divisualisasikan agar dapat membantu dalam proses pengambilan keputusan strategis yang sesuai dengan perspektif pelanggan.

Pada penelitian terdahulu oleh Marrese-Taylor et al. (2013) telah melakukan penelitian dengan menerapkan Analisis Sentimen (*Opinion Mining*) pada tingkat aspek dan menerapkannya pada domain pariwisata. Hasil dari penelitian memberikan kesimpulan bahwa ulasan mengenai produk pariwisata yang tersedia di *Web* memiliki informasi berharga mengenai preferensi wisatawan terkait produk pariwisata dan bisa dilakukan ekstraksi terhadapnya dengan menggunakan pendekatan Analisis Sentimen pada tingkatan aspek. Wijayanti dan Arisal (2017) telah dilakukan penelitian terkait Analisis Sentimen terhadap teks berbahasa Indonesia dengan tujuan mengevaluasi algoritme *Machine Learning* dalam menentukan polaritas dari teks. Fachrina dan Widyantoro (2017) telah dilakukan penelitian terkait Analisis Sentimen terhadap ulasan *online marketplace* dengan mengkombinasikan pendekatan *rule-based* dan *Machine Learning*. Dari kedua penelitian tersebut disimpulkan bahwa Analisis Sentimen pada tingkat aspek bisa diterapkan dalam domain pariwisata untuk mendapatkan preferensi wisatawan terkait produk pariwisata serta penggunaan SVM memberikan hasil yang paling optimal dalam Analisis Sentimen pada tingkat aspek. Oleh karena itu, pada penelitian ini juga melakukan Analisis Sentimen pada tingkat aspek dengan menggunakan SVM yang dikombinasikan dengan *Term Frequency-Inverse Document Frequency* (TF-IDF) sebagai *Term Weighting* yang bertujuan untuk mendapatkan sentimen pelanggan terhadap aspek pariwisata pantai Malang Selatan.

Berdasarkan latar belakang yang telah dijelaskan, penulis tertarik untuk melakukan penelitian untuk melakukan Analisis Sentimen terhadap opini berbahasa Indonesia pada *Web* yang diberikan oleh pelanggan tentang objek wisata pantai Malang Selatan. Akuisisi data opini bersumber dari situs TripAdvisor dengan memanfaatkan teknik *Web Scraping*. Analisis Sentimen dilakukan pada level aspek dengan menggunakan metode SVM dengan menggunakan TF-IDF sebagai *Term Weighting* dan ditampilkan melalui visualisasi *Dashboard*. Hasil visualisasi ini diharapkan dapat memudahkan pemangku kebijakan dalam memahami perspektif pelanggan serta mendapatkan wawasan yang mendukung untuk pengambilan keputusan strategis terkait aspek wisata pantai Malang Selatan. Oleh karena itu, penulis ingin mengangkat judul **“Analisis Sentimen Opini Pelanggan Terhadap Aspek Pariwisata Pantai Malang Selatan Menggunakan TF-IDF Dan Support Vector Machine”**

## 1.2 Rumusan masalah

Berdasarkan uraian latar belakang di atas, berikut merupakan rumusan permasalahan yang menjadi fokus dalam penelitian ini:

1. Bagaimana hasil penerapan teknik *Web Scraping* untuk mendapatkan data opini dari pelanggan terhadap aspek pariwisata pantai Malang Selatan yang bersumber dari situs TripAdvisor?

2. Bagaimana hasil penerapan *Support Vector Machine* dengan menggunakan TF-IDF sebagai *Term Weighting* untuk proses klasifikasi sentimen terhadap aspek pariwisata pantai Malang Selatan?
3. Bagaimana akurasi hasil klasifikasi sentimen menggunakan *Support Vector Machine* dengan TF-IDF terhadap setiap aspek pariwisata pantai Malang Selatan?
4. Bagaimana visualisasi *Dashboard* dan tingkat *Usability* dari *Dashboard* yang memuat hasil Analisis Sentimen opini pelanggan terhadap aspek dari pariwisata pantai Malang Selatan dapat membantu pemangku kebijakan dalam pengambilan keputusan strategis?

### 1.3 Tujuan

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Mendapatkan data opini dari pelanggan terhadap aspek pariwisata pantai Malang Selatan yang bersumber dari situs TripAdvisor.
2. Melakukan klasifikasi sentimen pada data opini dari pelanggan terhadap aspek dari pariwisata pantai Malang Selatan.
3. Melakukan uji akurasi hasil klasifikasi sentimen untuk mengukur tingkat ketepatan klasifikasi.
4. Menghasilkan visualisasi *Dashboard* dan tingkat *Usability* dari *Dashboard* yang memuat hasil Analisis Sentimen opini pelanggan terhadap aspek dari pariwisata pantai Malang Selatan yang dapat membantu pemangku kebijakan dalam pengambilan keputusan strategis.

### 1.4 Manfaat

Manfaat dari penelitian ini adalah:

1. Memberikan informasi terkait Analisis Sentimen opini pelanggan terhadap aspek dari pariwisata pantai Malang Selatan melalui hasil visualisasi *Dashboard*.
2. Menyajikan informasi yang dapat digunakan sebagai pendukung untuk pemangku kebijakan dalam pengambilan keputusan strategis terkait pengelolaan dan pengembangan pariwisata pantai Malang Selatan.

### 1.5 Batasan masalah

1. Penelitian dilakukan pada objek pariwisata pantai Malang Selatan.
2. Data bersumber dari situs TripAdvisor.
3. Analisis Sentimen dilakukan pada tingkatan aspek. Aspek yang dimiliki pariwisata pantai Malang Selatan meliputi aspek Umum, Kebersihan, Keramaian, Akses Jalan, dan kondisi Ombak.

4. Klasifikasi sentimen menggunakan algoritme *Support Vector Machine* (SVM) dengan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) sebagai *Term Weighting*.

## 1.6 Sistematika pembahasan

### Bab 1      **Pendahuluan**

Bab ini memuat hal-hal terkait latar belakang permasalahan, rumusan masalah, tujuan, manfaat, batasan masalah, dan sistematika pembahasan

### Bab 2      **Landasan Kepustakaan**

Bab ini memuat hal-hal terkait kajian pustaka dari penelitian terdahulu. Selain itu juga terdapat referensi dan dasar teori dalam melakukan Analisis Sentimen.

### Bab 3      **Metodologi**

Bab ini memuat hal-hal terkait langkah-langkah yang dilakukan dalam penelitian. Langkah-langkah penelitian meliputi identifikasi masalah, studi literatur, pengumpulan data, *Text Preprocessing*, *Term Weighting*, klasifikasi sentimen, analisis dan pengujian hasil klasifikasi, pembuatan visualisasi hasil Analisis Sentimen.

### Bab 4      **Perancangan**

Bab ini memuat tahap-tahap perancangan, juga termasuk di dalamnya hasil penerapan teknik *Web Scraping* untuk mendapatkan data opini pelanggan dari situs TripAdvisor, pemilihan aspek, pelabelan dataset, dan perancangan antar muka *Dashboard*.

### Bab 5      **Implementasi**

Bab ini memuat tentang tahap *Text Preprocessing* dalam proses *Text Mining*, *Term Weighting*, dan klasifikasi sentimen.

### Bab 6      **Analisis dan Pembahasan**

Bab ini memuat tentang hasil klasifikasi sentimen dari opini pelanggan terhadap setiap aspek wisata pantai Malang Selatan dengan menggunakan metode SVM. Dalam bab ini pula dibahas mengenai hasil pengujian akurasi terhadap hasil klasifikasi sentimen, serta visualisasi dari hasil Analisis Sentimen tersebut.

### Bab 7      **Penutup**

Bab ini memuat hal-hal terkait kesimpulan yang didapatkan dari hasil penelitian dan saran yang diberikan untuk perbaikan penelitian di masa mendatang.



## BAB 2 LANDASAN KEPUSTAKAAN

### 2.1 Kajian pustaka

Sub-bab ini membahas terkait penelitian-penelitian terdahulu yang pernah dilakukan antara lain oleh Fachrina dan Widyantoro (2017) telah melakukan penelitian mengenai klasifikasi *aspect* dan sentimen dari opini yang diberikan pengguna *online marketplace*. Pendekatan yang dilakukan menggunakan kombinasi antara *rule-based* dan *Machine Learning approach*. Klasifikasi opini dilakukan menggunakan algoritme *Naïve Bayes Classifier* dan *Support Vector Machine* (SVM). Tahapan penelitian meliputi *Sentence Preprocessing*, *Aspect and Sentiment classification*, *Opinion Sentiment Slassification*, dan terakhir *Aspect Based Summary*. *Sentence Preprocessing* berisi tahap-tahap mempersiapkan opini untuk dapat diproses untuk diklasifikasikan. Tahap-tahap tersebut antara lain penghapusan tanda baca dan tanda tanya, formalisasi kata ke dalam bentuk baku menggunakan pustaka InaNLP, koreksi ejaan menggunakan algoritme *Damerau-Levenshtein*, alih bahasa kata dari bahasa lain maupun khusus ke dalam Bahasa Indonesia, *Lemmatization* untuk menghilangkan imbuhan dan mendapatkan kata dasarnya, dan penghilangan *Stop words*. Klasifikasi *Aspect* dan sentimen dilakukan menggunakan *tool* Weka 3.7.10 dengan menerapkan algoritme SVM dan *Naïve Bayes Classifier*. Hasil dari penelitian tersebut SVM yang dikombinasikan dengan pendekatan *Rule-based* menunjukkan peforma terbaik dengan *F-Measure* berkisar antara 78,9% hingga 92%.

Penelitian lainnya dari Wijayanti dan Arisal (2017) yang membahas mengenai analisis polarisasi terhadap *User Generated Content* (UGC) yang menggunakan Bahasa Indonesia. Tujuan dari penelitian ini adalah melakukan evaluasi terhadap kombinasi dari *Feature Vectors* dan melakukan eksplorasi teknik *Ensemble* yang diperkenalkan Dietterich (2000) terhadap beberapa algoritme *Machine Learning* yang digunakan untuk melakukan klasifikasi. Beberapa metode *Feature Vectors* yang dikombinasikan antara lain *Term Frequency-Inverted Document Frequency* (TF-IDF), *Term Presence* (TP), *Sentiment Lexicon* (SL). Sedangkan algoritme *Machine Learning* yang digunakan untuk teknik *ensemble* antara lain SVM, *Naive Bayes*, *Logistic Regression*, dan *Decision Tree*. Hasil dari penelitian ini menunjukkan teknik *Ensemble* dapat meningkatkan akurasi dan *F1-Score*. Dalam penelitian ini juga menunjukkan TF-IDF yang dikombinasikan dengan SVM menghasilkan akurasi hingga 91,42% dan hanya berselisih sedikit dengan hasil dari penggunaan teknik *Ensemble* yang dikombinasikan dengan TF-IDF, TP, dan SL yang mendapatkan hasil 91,52%.

Marrese-Taylor et al. (2013) melakukan penelitian dengan menerapkan Analisis Sentimen (*Opinion Mining*) pada tingkat aspek dan menerapkannya pada domain pariwisata. Tujuan dari penelitian ini adalah mengajukan metode alternatif untuk mendapatkan preferensi dari wisatawan terhadap produk-produk pariwisata seperti hotel dan restoran. Data opini yang digunakan bersumber dari

ulasan yang ada di *Web*. Untuk mengevaluasi metode yang diajukan, dilakukan pula eksperimen menggunakan data ulasan restoran dan hotel yang bersumber dari situs TripAdvisor. Hasil dari penelitian memberikan kesimpulan bahwa ulasan mengenai produk pariwisata yang tersedia di *Web* memiliki informasi berharga mengenai preferensi wisatawan terkait produk pariwisata dan bisa dilakukan ekstraksi terhadapnya dengan menggunakan pendekatan Analisis Sentimen pada tingkatan aspek. Kemudian metode alternatif yang diajukan terbukti sangat efektif dalam menentukan orientasi sentimen dari suatu ulasan/opini, hal tersebut dibuktikan dengan hasil pengujian *Precision* dan *Recall* dengan nilai 90%. Kekurangan dari penelitian ini terletak pada rendahnya ekstraksi dari ekspresi aspek (*Aspect Expression*) yang berkisar pada 35% dari ekspresi aspek eksplisit yang ada pada data ulasan.

Ekawati dan Khodra (2017) melakukan penelitian terkait dengan sentimen analisis tingkat aspek dengan menerapkan model yang di pakai oleh Hercig et al. (2016). Analisis sentimen dilakukan terhadap ulasan restoran Indonesia dengan menerapkan 6 tahap yakni: *Text Preprocessing*, ekstraksi aspek, kategorisasi aspek, klasifikasi sentimen, pembuatan struktur opini, dan kalkulasi pemeringkatan (*rating*). Jumlah data yang digunakan adalah 992 kalimat yang di pergunakan sebagai bahan eksperimen untuk proses menemukan kombinasi *feature* yang terbaik untuk proses ekstraksi serta kategorisasi aspek, dan proses klasifikasi sentimen terhadap tiap aspek. Hasil pengujian dari proses ekstraksi aspek, kategorisasi aspek, dan klasifikasi sentimen memperoleh nilai F1-measure masing-masing sebesar 0.793, 0.823, 0.642.

Bucur (2015) melakukan penelitian yang bertujuan untuk menerapkan *Opinion Mining* untuk melakukan ekstraksi dan klasifikasi ulasan hotel yang diberikan oleh pengguna dalam situs pariwisata. Proses klasifikasi menghasilkan hasil klasifikasi ke dalam tiga kategori yakni positif, negatif, dan netral. Hasil dari penelitian ini memberikan kesimpulan bahwa proses yang telah di jalankan menghasilkan tingkat akurasi yang bisa diterima dan memberikan keuntungan berupa karakter domain independent serta tidak membutuhkan banyak sumber daya untuk melakukannya. Kesimpulan selanjutnya adalah dalam domain pariwisata analisis yang berorientasi pada tingkat aspek akan meningkatkan hasil dari proses Analisis Sentimen, hal tersebut dikarenakan beragamnya aspek yang diulas oleh pengguna serta variasi sentimen yang di ekspresikan pula oleh pengguna dalam ulasannya.

Nathania et al. (2018) melakukan penelitian terkait permasalahan klasifikasi terhadap data yang bersumber dari Media Sosial Twitter. Tujuan dari penelitian ini adalah melakukan klasifikasi *Tweet* untuk mengetahui jenis kategori *Tweet Spam* atau bukan *Spam*. Tahapan penelitian meliputi *Preprocessing*, yakni *Cleansing*, *Case Folding*, tokenisasi, *Filtering*, dan *stemming*. Dilanjutkan dengan proses *Term Weighting*, hingga proses klasifikasi dengan menggunakan metode *Improved K-Nearest Neighbor*. Penelitian ini menghasilkan rata-rata *Precision*

sebesar 0.8946, *Recall* sebesar 0.9405, *F-Measure* sebesar 0.9155 dan hasil akurasi sebesar 89.57%.

Dari beberapa penelitian tersebut, pada skripsi ini membahas mengenai Analisis Sentimen pada tingkat aspek yang dilakukan dengan menggunakan pendekatan *Machine Learning* menggunakan algoritme *Support Vector Machine* yang dikombinasikan dengan TF-IDF sebagai metode *Term Weighting*. Perbedaan penelitian ini dengan penelitian terdahulu yakni pada data opini pelanggan yang merupakan ulasan terhadap objek pariwisata pantai Malang Selatan dengan aspeknya yakni aspek Umum, Kebersihan, Keramaian, Akses Jalan, dan kondisi Ombak.

## 2.2 Profil Dinas Pariwisata dan Kebudayaan Kabupaten Malang

Dinas Pariwisata dan Kebudayaan (Disparbud) Kabupaten Malang adalah unsur pelaksana otonomi daerah di bidang kebudayaan dan pariwisata Kabupaten Malang. Disparbud Kabupaten Malang beralamat di Jl. Raya Singosari No.275. Disparbud Kabupaten Malang saat ini dipimpin oleh Bapak Made Arya Wedanthara, SH, M.Si. Disparbud Kabupaten Malang memiliki visi “Terwujudnya Kepariwisataan Kabupaten Malang Yang Berbasis Masyarakat”. Dengan misi sebagai berikut:

1. Membangun jati diri dan citra kepariwisataan Kabupaten Malang yang berbasis masyarakat.
2. Mendorong perkembangan kepariwisataan Kabupaten Malang yang berkualitas dan memiliki daya saing melalui:
  - a. Pengembangan objek dan daya tarik wisata yang berdasarkan kearifan lokal;
  - b. Membangun sarana dan prasarana dalam keselarasan dan keharmonisan lingkungan;
  - c. Mewujudkan kualitas pelayanan yang baik pada masyarakat;
  - d. Mengoptimalkan sarana informasi dan menyelenggarakan promosi yang lebih berkualitas.
3. Meningkatkan peran serta masyarakat.

### 2.2.1 Pengawasan dan evaluasi kegiatan pengelolaan pariwisata

Dinas Pariwisata dan Kebudayaan (Disparbud) Kabupaten Malang bertindak sebagai pembina dari Perusahaan Daerah (PD) Jasa Yasa. PD Jasa Yasa merupakan perusahaan daerah milik Pemerintah Daerah Kabupaten Malang yang memiliki wewenang dalam pengelolaan unit usaha, objek wisata pantai Malang Selatan masuk ke dalam kategori tersebut. Sebagai pembina PD Jasa Yasa, Disparbud Kabupaten Malang memiliki kewajiban untuk melakukan kegiatan pengawasan dan evaluasi, terutama terhadap kegiatan pengelolaan dan pengembangan objek pariwisata. Berdasarkan hasil wawancara yang dilakukan dengan pegawai terkait dari Disparbud Kabupaten Malang, menyatakan bahwa

tidak ada prosedur khusus dalam proses pengawasan dan evaluasi. Proses pengawasan dilakukan dengan cara observasi langsung ketika ada kegiatan tahunan atau kegiatan besar dilakukan pada suatu objek wisata. Hasil temuan observasi tersebut kemudian dijadikan bahan evaluasi yang kemudian dikomunikasikan dengan PD Jasa Yasa untuk kemudian dijadikan pertimbangan dalam kegiatan pengelolaan objek pariwisata.

## 2.3 Opini

### 2.3.1 Teori Opini

Moore (1988) mengatakan bahwa opini memiliki pengertian sebagai suatu kesimpulan yang ada di dalam pemikiran dan belum dikeluarkan untuk dapat dijadikan bahan untuk diperdebatkan. Suatu opini yang sudah bersifat tetap dipandang sebagai sentimen, bila dipegang secara teguh maka dianggap sebagai keyakinan, sementara bila dipengaruhi oleh kecenderungan maka dinamakan pandangan. Sedangkan menurut Soenarjo (1997) opini merupakan jawaban terbuka (*overt*) terhadap suatu persoalan atau isu ataupun jawaban yang diajukan secara tertulis maupun lisan. Opini juga dianggap sebagai jawaban lisan pada individu yang memberi respon atau tanggapan kepada rangsangan di mana suatu situasi atau keadaan yang pada umumnya diajukan suatu pertanyaan. Berdasarkan definisi di atas, opini adalah pendapat atau pernyataan seseorang yang dinyatakan baik melalui kata-kata, tulisan, tindakan, atau cara-cara lain yang mengandung arti.

### 2.3.2 Komponen Opini

Menurut Rivers dan Mathews (1994) pada dasarnya setiap opini memiliki 3 komponen.

1. Kepercayaan

Kepercayaan mengacu pada sesuatu yang diterima khalayak, benar atau tidak berdasarkan pengalaman masa lalu, pengetahuan dan informasi sekarang serta persepsi yang berkesinambungan.

2. Nilai

Nilai melibatkan kesukaan-ketidaksukaan, cinta dan kebencian, hasrat dan ketakutan, bagaimana orang menilai sesuatu dan intensitas penilaiannya apakah kuat, lemah, netral.

3. Pengharapan

Mengandung citra seseorang tentang apa keadaannya setelah dilakukannya tindakan. Pengharapan, ditentukan dari pertimbangan terhadap sesuatu yang terjadi pada masa lalu, keadaan sekarang, dan sesuatu yang kira-kira akan terjadi jika dilakukan perbuatan tertentu.

## 2.4 Web Scraping

Mitchell (2018) menjelaskan bahwa yang dimaksudkan sebagai *Web Scraping* adalah kegiatan pengumpulan (*gathering*) data yang bersumber dari Internet. Dalam praktiknya, *Web Scraping* adalah melakukan pengumpulan data dengan memanfaatkan *Application Programming Interface* (API) atau secara umum sama seperti interaksi manusia dengan halaman *Web* melalui *Web browser* tapi perbedaannya terletak pada tujuannya. Tujuan dari *Web Scraping* adalah mendapatkan data untuk kemudian melakukan ekstraksi informasi yang dimiliki oleh data tersebut. *Web Scraping* memiliki karakteristik bekerja secara cepat, efisien, berjalan secara otomatis, dan menghasilkan data dalam format yang lebih terstruktur dan lebih mudah digunakan (Castrillo-Fernandez, 2015). Cara kerja *Web Scraping* adalah dengan mengakses halaman *Web*, memilih elemen data yang ada dalam halaman tersebut, melakukan ekstraksi dan transformasi bila diperlukan, dan terakhir menyimpan data tersebut menjadi *dataset* terstruktur (Boeing dan Waddell, 2017).

### 2.4.1 WebHarvy

SysNucleus (2017) sebagai pengembang WebHarvy mengatakan bahwa WebHarvy adalah sebuah *Visual Web Scraper* yang memudahkan untuk melakukan *Scrape* terhadap data terstruktur dari banyak halaman *Web*. WebHarvy memiliki kelebihan sebagai berikut:

1. *Scraping* data menggunakan *point and click Interface*

Untuk dapat melakukan *Scraping* cukup dengan menggunakan antarmuka *point and click* yang berarti cukup telusuri ke halaman yang berisi data yang akan diekstrak dan pilih data yang diperlukan dengan mengkliknya.

2. Identifikasi pola otomatis

Jadi ketika proses *Scraping* bertujuan untuk mendapatkan data yang berulang maka WebHarvy tidak akan menanyakan apakah mencoba untuk menangkap satu item data atau seluruh daftar item. Jika data berulang, WebHarvy akan secara cerdas mengidentifikasinya.

3. Melakukan *Scraping* terhadap *list* dengan multi kategori

WebHarvy memungkinkan untuk melakukan *Web Scraping* terhadap data yang berbentuk *list* yang ada di beberapa kategori pada suatu situs *Web* menggunakan satu konfigurasi.

4. *Scraping* berdasarkan kata kunci

Fitur *Scraping* dengan kata kunci membantu untuk secara otomatis melakukan pencarian menggunakan daftar kata kunci dan melakukan *Scraping* terhadap data yang dihasilkan, untuk setiap kata kunci.



5. Fitur-fitur tambahan

Dalam WebHarvy terdapat pilihan untuk menangkap data dari kode HTML dari bagian yang dipilih pada halaman *Web*, bisa juga menerapkan *Regular Expression (RegEx)* pada teks atau HTML untuk melakukan pemilihan data dengan lebih akurat, ada pula fitur untuk menjalankan kode *JavaScript* pada halaman sebelum memulai *Scraping* dan fitur lainnya.

6. *Scraping* dengan anonim

*Scraping* anonim dilakukan dengan memanfaatkan *Server Proxy* untuk mencegah target *Web Server* memblokir alamat IP komputer lokal.

7. Penyimpanan ke berbagai format

WebHarvy memberikan opsi untuk menyimpan data hasil proses *Scraping* kedalam beberapa format antara lain CSV (*Comma Separated Values*), TSV (*Tab Separated Values*), XML (*Extensible Markup Language*), maupun JSON (*JavaScript Object Notation*). Selain diubah menjadi file, WebHarvy juga menyediakan pilihan untuk menyimpan data hasil proses *Scraping* ke basis data SQL.

## 2.5 Data Mining

### 2.5.1 Definisi Data Mining

*Data Mining* merupakan satu bagian dari proses *Knowledge Discovery*. *Data Mining* adalah proses pencarian pola dan pengetahuan dari sumber data yang besar. Sumber data yang besar yang di maksud bisa merupakan basis data, *Data Warehouse*, *Web*, atau sumber penyimpanan informasi lain, atau bisa juga berupa data yang dimasukkan ke dalam sistem *Data Mining* secara dinamis. (Han et al., 2011).

### 2.5.2 Tahapan dalam Data Mining

Liu (2013) menjelaskan bahwa untuk menerapkan *Data Mining* perlu untuk memahami pada domain apa *Data Mining* akan diterapkan. Untuk dapat mengetahui hal tersebut diperlukan analisis yang dilakukan oleh *Data Analyst*, tahapan analisis tersebut bertujuan untuk memahami kebutuhan data, serta sumber data yang bisa digunakan untuk memenuhi kebutuhan tersebut. Hal tersebut mutlak dilakukan agar *Data Mining* dapat diterapkan. Penerapan *Data Mining* dibagi menjadi 3 tahap yakni:

1. *Pre-processing*

Data mentah yang akan diproses dalam *Data Mining* biasanya tidak sesuai untuk *Mining* dikarenakan beberapa alasan misalnya terdapat gangguan dalam data itu sendiri contohnya adanya *noise* yang perlu dihilangkan atau abnormalitas yang perlu dinormalkan. Selain adanya gangguan semacam itu, ukuran data turut juga mempengaruhi. Data yang terlalu besar bisa menjadi masalah karena ada adanya atribut yang tidak



relevan namun masuk dalam tahap *Data Mining*. Inti dari tahap *Pre-processing* ini adalah pengkondisian data untuk selanjutnya diolah dalam proses *Data Mining*.

## 2. *Data Mining*

Tahap ini adalah tahap inti dimana data yang telah dikondisikan sebelumnya diproses dengan menggunakan algoritme *Data Mining* untuk menghasilkan pola ataupun informasi.

## 3. *Post-processing*

Pada tahap ini merupakan tahap untuk melakukan evaluasi terhadap hasil dari *Data Mining* yang telah dilakukan. *Data Mining* belum tentu menghasilkan pola maupun informasi yang berguna, sehingga pada tahap ini bertujuan untuk melakukan identifikasi terhadap hasil *Data Mining* yang bermanfaat dan sesuai dengan tujuan pengaplikasian *Data Mining*. Teknik yang biasa dipakai dalam tahap ini adalah evaluasi dan visualisasi.

### 2.5.3 Teknik analisis data dalam *Data Mining*

Sharma et al. (2017) menerangkan beberapa teknik analisis data yang bisa dilakukan dalam *Data Mining*:

#### a. *Characterization*

Bertujuan untuk melakukan generalisasi, perangkuman terhadap karakteristik data yang berbeda. Dalam *Characterization* dilakukan peringkasan karakteristik umum atau fitur dari kelas target data. Data yang sesuai dengan kelas yang ditentukan pengguna biasanya dikumpulkan menggunakan *Query*.

#### b. *Classification*

Proses yang bertujuan untuk melakukan klasifikasi terhadap data untuk diklasifikasikan ke dalam kelas-kelas yang berbeda sesuai dengan model klasifikasi yang digunakan.

#### c. *Regression*

Regresi memiliki kemiripan dengan klasifikasi namun perbedaannya terletak pada objek yang merupakan data kontinyu. Bila dibandingkan dengan klasifikasi maka regresi berfokus pada memprediksi nilai data numerik yang hilang atau tidak ada.

#### d. *Association*

Bertujuan untuk mendapatkan asosiasi antar objek. Objek yang dimaksud berupa sumber data yang dijadikan sumber dalam proses *Data Mining*. Hasil teknik ini bisa berupa asosiasi antara berbagai basis data atau juga antar atribut yang ada dalam sebuah basis data.

e. *Clustering*

*Clustering* adalah melakukan pengelompokan data ke dalam beberapa kelas baru yang mampu merepresentasikan data tersebut. Teknik ini memecah data yang besar ke dalam kelompok yang lebih kecil dengan tujuan untuk memudahkan proses desain dan implementasi. Dalam teknik ini bertujuan untuk memaksimalkan kemiripan antara objek dalam suatu kelas, dan mengurangi kemiripan antar kelas yang ada.

f. *Change Detection*

Teknik ini bertujuan untuk melakukan identifikasi terhadap perubahan yang terjadi pada data dengan cara membandingkannya dengan data sebelumnya.

g. *Deviation Detection*

Teknik ini memiliki konsen pada deviasi yang terbentuk antara nilai sebenarnya dengan nilai yang diperkirakan dari objek. Teknik ini bekerja berdasarkan variabel waktu dan atau perbedaan antar *subset* data.

h. *Link Analysis*

Teknik ini menggunakan teori graf untuk menemukan hubungan antar objek dengan tujuan membentuk model berdasarkan pola yang ada dalam hubungan yang ditemukan.

i. *Sequential Pattern Mining*

Mabroukeh dan Ezeife (2010) menjelaskan bahwa teknik ini berfokus pada menemukan pola yang relevan secara statistik yang berasal dari data yang nilai-nilainya berurutan (*sequential*).

#### 2.5.4 Text Mining

*Text Mining* adalah salah area studi yang menggabungkan teknik dari *Data Mining*, *Machine Learning*, *Natural Language Processing* (NLP), *Information Retrieval* serta *Knowledge Management*. Tujuan dari *Text Mining* tidak jauh berbeda dengan *Data Mining* yang tidak lain adalah menemukan informasi/pengetahuan yang berguna dari sumber data yang besar dengan cara mengidentifikasi dan melakukan eksplorasi terhadap pola yang menarik dari sumber data. Adapun perbedaan mendasar antara keduanya adalah terletak pada bentuk dari data yang dijadikan sumber dalam proses pencarian *knowledge*. Dalam *Data Mining* bentuk data yang digunakan berupa data terstruktur yang bersumber dari basis data, *Data Warehouse*, maupun repositori lainnya, sedangkan dalam *Text Mining* menggunakan data tidak terstruktur dalam bentuk teks (Feldman & Sanger, 2007).

##### 2.5.4.1 Arsitektur Text Mining

Feldman & Sanger (2007) mendefinisikan bahwa dalam tingkatan fungsional, *Text Mining* mengadopsi model umum yang biasa dipakai dalam

penerapan *Data Mining*. Model tersebut secara garis besar terbagi dalam empat area atau tahapan yakni:

a. *Preprocessing tasks*

Dalam tahap ini termasuk di dalamnya semua kegiatan, proses, dan metode yang dipergunakan untuk mempersiapkan data agar sesuai untuk tahapan inti *Text Mining*. Tahapan inti yang dimaksud adalah operasi untuk menemukan pengetahuan/informasi. Secara umum tahap ini melakukan konversi informasi dari setiap sumber data ke dalam bentuk yang sesuai sebelum melakukan berbagai metode ekstraksi fitur terhadapnya. Beberapa contoh tugas dalam tahapan ini antara lain:

1. *Case folding* yang bertujuan untuk mengubah seluruh huruf menjadi bentuk huruf kecil (*lowercase*).
2. *Cleansing* yakni proses yang bertujuan untuk menghilangkan elemen dalam teks yang dianggap sebagai *noise*.
3. *Tokenizing* atau tokenisasi dilakukan untuk memisahkan semua kata yang dipisahkan oleh spasi atau *whitespace* dalam teks.
4. *Stemming* adalah proses yang bertujuan untuk mengubah kata ke dalam bentuk dasarnya atau kata dasarnya.
5. *Stop-word removal* bisa disebut juga atau termasuk dalam proses *filtering*, tujuannya adalah menghilangkan kata-kata yang masuk ke dalam kategori *stop-word*.

b. *Core mining operations*

Tahap ini merupakan inti dari *Text Mining* yakni melakukan pencarian pola, analisis tren, dan penerapan algoritme pencarian informasi secara inkremental. Tahap ini biasa disebut juga sebagai proses *Knowledge Distillation*.

c. *Presentation layer components*

Bagian ini termasuk di dalamnya berupa *Graphical User Interface* (GUI) dan fungsi penelusuran pola, serta bisa juga akses menggunakan *Query*.

d. *Refinement Techniques*

Tahap ini biasa disebut juga *postprocessing* dengan tujuan optimasi dalam proses pencarian untuk menemukan pola. Beberapa pendekatan yang biasa digunakan dalam tahap ini diantaranya *Suppression*, *Ordering*, *Pruning*, *Generalization*, dan *Clustering*.

## 2.6 Analisis Sentimen

Analisis sentimen atau sering disebut juga *Opinion Mining* adalah bidang studi yang menganalisis opini, sentimen, evaluasi, perilaku, dan emosi yang tertuju kepada entitas tertentu seperti produk, layanan, organisasi, dan sebagainya (Liu, 2012). Terdapat 3 tingkatan analisis pada Analisis Sentimen yakni:

### 1. Level dokumen

Tugas dalam tingkatan ini adalah melakukan klasifikasi terhadap opini dalam sebuah dokumen. Opini dalam dokumen tersebut dianalisis lalu diklasifikasikan menjadi sentimen positif atau sentimen negatif (Pang et al., 2002; Turney, 2002). Tingkatan analisis ini memegang asumsi bahwa setiap dokumen berisikan opini terhadap sebuah entitas tertentu saja. Sehingga tidak memungkinkan bila dokumen tersebut berisi opini terhadap lebih dari satu entitas.

### 2. Level kalimat

Pada tingkatan ini, kalimat yang dijadikan objek untuk diklasifikasikan sentimennya. Sentimen pada level ini dikelompokkan menjadi positif, negatif, dan netral yang berarti tidak berisikan opini. Tujuan dari level ini hampir sama dengan *Subjectivity Classification*. *Subjectivity Classification* memisahkan antara kalimat yang mengandung fakta dengan kalimat yang mengandung opini (wiebe et al., 1999).

### 3. Level entitas dan aspek

Analisis pada tingkatan dokumen dan kalimat tidak dapat menemukan apa sebenarnya yang disukai dan tidak disukai dari sebuah entitas yang dijadikan objek opini. Dengan melakukan analisis pada tingkatan aspek, maka yang dijadikan objek analisis adalah opini itu sendiri sehingga hasil analisis yang didapatkan akan lebih baik. Hal tersebut bersumber dari asumsi bahwa sebuah opini memiliki 2 komponen yakni sentimen (positif/negatif) dan target dari opini tersebut.

## 2.6.2 Kerangka kerja Analisis Sentimen

Liu (2012) mendefinisikan tujuan dari Analisis Sentimen adalah menemukan kuintupel  $(e_i, a_{ij}, s_{jkl}, h_k, t_l)$  dalam sebuah opini. Kuintupel tersebut terdiri dari nama dari entitas  $(e_i)$ , aspek dari entitas  $e_i$   $(a_{ij})$ , sentimen terhadap aspek  $a_{ij}$  dari entitas  $e_i$   $(s_{jkl})$ , pemegang opini  $(h_k)$ , dan waktu opini tersebut diungkapkan oleh pemegang opini  $(t_l)$ .

Berikut adalah tahapan dalam melakukan Analisis Sentimen:

#### 1. Ekstraksi dan pengkategorian entitas

Melakukan ekstraksi terhadap semua entitas yang ada dalam sebuah dokumen berisi opini, lalu mengelompokkannya kedalam berbagai kategori yang representatif.

#### 2. Ekstraksi dan pengkategorian aspek

Melakukan ekstraksi terhadap semua aspek yang dimiliki oleh entitas, lalu mengelompokkannya kedalam beberapa *cluster*.

3. Ekstraksi dan kategorisasi pemegang opini  
Melakukan ekstraksi terhadap semua pemegang opini yang ada dalam dokumen berisi opini atau data terstruktur, lalu mengelompokkannya kedalam kategori.
4. Ekstraksi dan penyeragaman waktu  
Melakukan ekstraksi waktu terkait kapan opini tersebut diberikan, dan melakukan standarisasi atau penyeragaman terhadap format waktu.
5. Klasifikasi sentimen terhadap aspek  
Menentukan opini terhadap suatu aspek dari suatu entitas ke dalam kategori positif, negatif atau netral, atau menggunakan angka dalam bentuk *rating* terhadap aspek.
6. Menghasilkan kuintupel opini  
Mendefinisikan semua kuintupel opini  $(e_i, a_{ij}, s_{jkl}, h_k, t_l)$  yang ada dalam suatu dokumen berisi opini.

## 2.7 Term Weighting

### 2.7.1 Term Frequency-Inverse Document Frequency (TF-IDF)

Manning et al. (2009) menjelaskan bahwa TF-IDF adalah kombinasi dari pembobotan *Term Frequency* dengan *Inverse Document Frequency*. *Term frequency* adalah pemberian bobot pada tiap *Term* atau istilah yang muncul pada suatu dokumen. Bobot tersebut bergantung pada jumlah munculnya istilah tersebut dalam dokumen. *Term Frequency* merupakan skema pembobotan dengan memberikan bobot pada suatu istilah  $t$  dalam dokumen  $d$  sesuai dengan banyaknya kemunculan  $t$  dalam  $d$ . *Term Frequency* memiliki notasi  $tf_{t,d}$ . *Document Frequency* adalah banyaknya dokumen yang mengandung istilah  $t$  sehingga dinotasikan  $df_t$ . *Document Frequency* digunakan untuk menghitung *Inverse Document Frequency* ( $idf$ ) dari istilah  $t$ . Berikut adalah rumus menghitung  $idf$  dari sebuah istilah  $t$ , dengan  $N$  sebagai banyaknya dokumen:

$$idf_t = \log \frac{N}{df_t} \quad (2.1)$$

Dimana:

- $idf_t$  : Nilai  $idf$  dari istilah  $t$
- $N$  : Banyaknya dokumen atau koleksi yang ada
- $df_t$  : Banyaknya kemunculan istilah  $t$  dalam dokumen  $d$

Selanjutnya pembobotan TF-IDF dilakukan dengan mengkombinasikan hasil pembobotan  $tf$  dan  $idf$  yang telah dilakukan untuk mendapatkan bobot gabungan untuk tiap istilah pada tiap dokumen. Berikut persamaan 2.2 adalah rumus menghitung TF-IDF dari sebuah istilah  $t$  dalam dokumen  $d$ :



$$tf-idf_{t,d} = tf_{t,d} \times idf_t \quad (2.2)$$

Dimana:

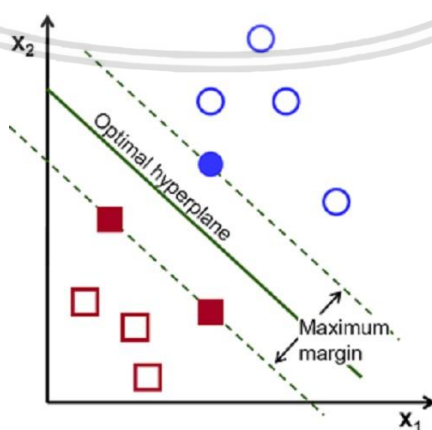
- $tf-idf_{t,d}$  : bobot TF-IDF dari istilah  $t$  dalam dokumen  $d$
- $tf_{t,d}$  : frekuensi munculnya istilah  $t$  dalam dokumen  $d$
- $idf_t$  : Nilai  $idf$  dari istilah  $t$

## 2.8 Machine Learning

*Machine learning* berarti menerapkan pemrograman komputer dengan tujuan untuk melakukan optimasi performa menggunakan data contoh atau pengalaman yang lalu. Model di definisikan sampai kepada parameter tertentu, lalu dilakukan pembelajaran (*learning*) dengan cara melakukan eksekusi terhadap program komputer untuk melakukan optimasi terhadap parameter tersebut dengan menggunakan data latih atau pengalaman yang lalu. Model yang digunakan bisa bersifat prediktif untuk melakukan prediksi masa depan, atau deskriptif untuk mendapatkan pengetahuan dari data, atau gabungan keduanya (Alpaydin, 2010).

### 2.8.1 Support Vector Machine (SVM)

Han et al. (2012) menjelaskan bahwa *Support Vector Machine* adalah metode klasifikasi untuk mencari *Maximum Marginal Hyperplane* (MMH) atau pemisah terbaik untuk menciptakan pemisahan maksimal untuk semua kelas. *Margin* bisa di definisikan sebagai jarak terpendek dari sebuah *hyperplane* terhadap satu sisi dari *margin* itu sama dengan jarak *hyperplane* dengan sisi *margin* lainnya, dengan catatan kedua *margin* tersebut dalam posisi paralel dengan *hyperplane*. Seperti yang ditunjukkan dalam Gambar 2.1. Kemudian dapat disimpulkan bahwa semakin besar *margin* bisa dikatakan memiliki akurasi yang lebih tinggi.



Gambar 2.1 Optimal Hyperplane



Dalam SVM menggunakan *set* data latih dalam bentuk  $(X_i, y_i)$   $X_i$  adalah *tuple* dan  $y_i$  adalah label kelas dengan  $i=1....N$ ,  $X_i \in R^d$  dan  $y_i \in \{-1, 1\}$ . Tujuan dari SVM adalah untuk dapat membentuk *classifier* seperti pada Persamaan 2.3 sebagai berikut.

$$f(x_i) = \begin{cases} \geq 0, & y_i = +1 \\ < 0, & y_i = -1 \end{cases} \quad (2.3)$$

Untuk membentuk *hyperplane* dijelaskan dalam Persamaan 2.4 sebagai berikut.

$$W \cdot X + b = 0 \quad (2.4)$$

Dimana :

- $W$  : Bobot vektor  $\{w_1, w_2, w_3, ..., w_n\}$   $n$  adalah banyaknya atribut
- $b$  : Nilai skalar, atau biasa disebut sebagai bias
- $X$  : Potongan data latih atau *training tuples*

Sehingga apabila  $b$  dianggap sebagai bobot tambahan maka  $b$  di definisikan sebagai  $w_0$ . Sehingga Persamaan 2.4 dapat dibentuk ulang menjadi Persamaan 2.5 berikut.

$$w_0 + w_1x_1 + w_2x_2 = 0 \quad (2.5)$$

Sehingga untuk semua titik yang berada pada sisi atas dari *hyperplane* maka persamaannya sesuai dengan Persamaan 2.6 berikut.

$$w_0 + w_1x_1 + w_2x_2 > 0 \quad (2.6)$$

Sedangkan untuk semua titik yang berada pada sisi bawah dari *hyperplane* maka persamaannya sesuai dengan Persamaan 2.7 berikut.

$$w_0 + w_1x_1 + w_2x_2 < 0 \quad (2.7)$$

Karena bobot bisa disesuaikan maka *hyperplane* yang membedakan sisi *margin* dapat dirumuskan sesuai dengan Persamaan 2.8 dan 2.9.

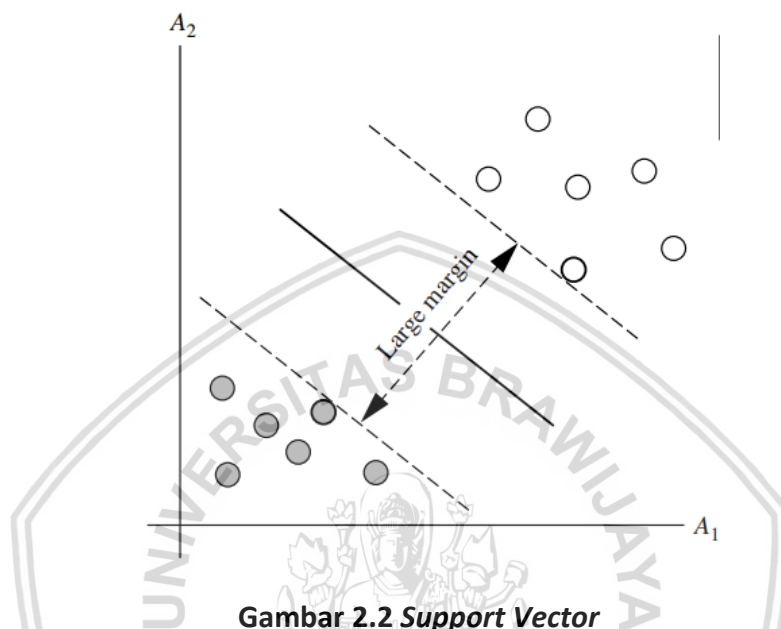
$$H_1: w_0 + w_1x_1 + w_2x_2 \geq 1 \quad \text{untuk } y_i = +1 \quad (2.8)$$

$$H_2: w_0 + w_1x_1 + w_2x_2 \leq -1 \quad \text{untuk } y_i = -1 \quad (2.9)$$

Persamaan diatas dapat mengartikan bahwa setiap *tuple* yang berada di atas atau tepat pada  $H_1$  maka masuk ke dalam kelas +1 begitu juga sebaliknya. Bila Persamaan 2.8 dan 2.9 dikombinasikan maka akan mendapatkan Persamaan 2.10

$$y_i(w_0 + w_1x_1 + w_2x_2) \geq 1, \forall i \quad (2.10)$$

Jadi untuk setiap *training tuple* yang berada tepat pada  $H_1$  atau  $H_2$  serta memenuhi Persamaan 2.9 maka *tuple* itu yang dinamakan sebagai *Support Vector*. Seperti pada Gambar 2.2 menunjukkan *Support Vector* yang sangat dekat dengan MMH. *Support Vector* adalah *tuple* yang paling sulit untuk di klasifikasikan namun sisi lainnya *Support Vector* juga memberikan informasi yang paling tinggi untuk proses klasifikasi.



**Gambar 2.2 Support Vector**  
Sumber : Han et al. (2012)

Adapun perhitungan MMH dapat dijadikan kedalam bentuk *Decision Boundary* yang ditunjukkan oleh Persamaan 2.10

$$d(\mathbf{X}^T) = \sum_{i=1}^l y_i a_i \mathbf{X}_i \mathbf{X}^T + b_0, \quad (2.11)$$

Dimana:

$y_i$  : Label kelas dari *Support Vector*  $X_i$

$X^T$  : *Test tuple*

$l$  : jumlah *Support Vector*

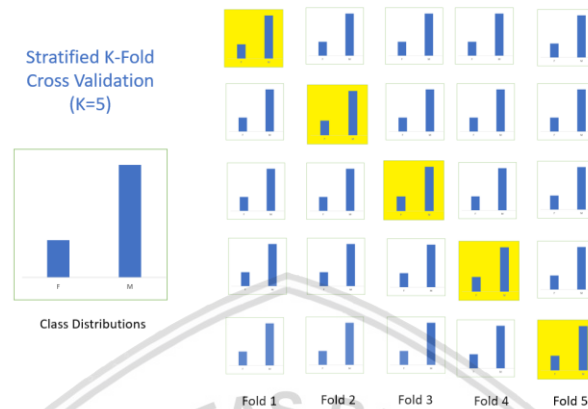
$a_i$  dan  $b_0$ : Parameter numerik yang ditentukan oleh optimasi atau SVM

## 2.9 Pemilihan dan evaluasi model

### 2.9.1 Kfold Cross-Validation

Han et al. (2012) Menjelaskan bahwa dalam *Kfold Cross Validation*, *dataset* awal dibagi secara acak menjadi sebanyak  $k$  bagian/*folds* dengan jumlah yang relatif sama untuk setiap bagian. *Folds* dinotasikan sebagai  $D_1, D_2, \dots, D_k$ . Kemudian tahapan *training* dan *testing* dilakukan sebanyak  $k$ . Dalam iterasi  $i$ , *fold*  $D_i$  dijadikan sebagai *test set*, dan *fold* lainnya digunakan untuk melakukan *train* model. *Stratified Kfold Cross Validation* melakukan distribusi data untuk setiap

kelas dengan tujuan untuk memastikan bahwa setiap *fold* memiliki rasio perbandingan yang sama dengan data awal. Gambar 2.3 menunjukkan contoh pembagian kelas untuk masing *fold* tetap sama dengan rasio awal *dataset*. Gambar 2.3 juga menunjukkan proses kombinasi *training* dan *testing* terhadap tiap *fold*.



**Gambar 2.3 Ilustrasi *Stratified 5 fold Cross Validation***

## 2.9.2 Metriks untuk evaluasi model

Han et al. (2012) menyebutkan ada 4 *building blocks* yang banyak digunakan sebagai bahan dalam banyak pengukuran evaluasi, yakni:

1. *True Positives* (TP): Data positif yang berhasil dilabeli dengan benar oleh *classifier*.
2. *True Negatives* (TN): Data negatif yang berhasil dilabeli dengan benar oleh *classifier*.
3. *False Positives* (FP): Data negatif yang ternyata salah dilabeli sebagai positif oleh *classifier*.
4. *False Negatives* (FN): Data positif yang ternyata salah dilabeli sebagai negatif oleh *classifier*.

Untuk lebih mudahnya keempat komponen tadi ditampilkan dalam *Confusion Matrix* seperti pada Tabel 2.1.

**Tabel 2.1 *Confusion Matrix***

		Kelas prediksi		Total
		TRUE	FALSE	
Kelas Sebenarnya	TRUE	TP	FN	P
	FALSE	FP	TN	N
Total		P'	N'	P + N

Sumber: Diadaptasi dari Han et al. (2012)

Metrik pengukuran yang biasa digunakan untuk mengevaluasi *classifier* adalah *Accuracy*. *Accuracy* menunjukkan persentase keberhasilan *classifier* dalam menentukan label. Berikut merupakan persamaan untuk menentukan *Accuracy*:

$$Accuracy = \frac{TP+TN}{P+N} \quad (2.3)$$

Metrik selanjutnya adalah *Precision* dan *Recall*. *Precision* bisa dianggap sebagai pengukuran untuk mendapatkan ketepatan (*exactness*), sedangkan *Recall* adalah pengukuran untuk mendapatkan *completeness*. Berikut adalah persamaan yang digunakan untuk mengukur *Precision* dan *Recall*:

$$Precision = \frac{TP}{TP+FP} \quad (2.4)$$

$$Recall = \frac{TP}{TP+FN} = \frac{TP}{P} \quad (2.5)$$

*Precision* dan *Recall* juga dapat digunakan untuk membentuk metrik pengukuran lainnya yakni *F measure* atau *F<sub>1</sub> Score* atau *F-Score*. *F<sub>1</sub> Score* adalah hasil perhitungan yang merupakan *harmonic mean* dari *Precision* dan *Recall*. Berikut adalah persamaan yang digunakan untuk mengukur *F<sub>1</sub> Score*:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.6)$$

## 2.10 Bahasa Pemrograman Python

Python diciptakan oleh Guido van Rossum di Belanda pada tahun 1990. kemudian Python menjadi Bahasa Pemrograman yang dipakai secara luas dalam industri dan pendidikan karena sederhana, ringkas, sintaksis yang intuitif dan memiliki pustaka yang luas (Liang, 2013). Sedangkan menurut Python Software Foundation (2018) selaku pengembang Python mendefinisikannya sebagai bahasa yang mudah dipelajari, Bahasa Pemrograman yang *powerful*. Python memiliki struktur data tingkat tinggi yang efisien dan pendekatan yang sederhana namun efektif untuk pemrograman berorientasi objek. Sintaksis dan pengetikan dinamis Python yang elegan, bersama dengan sifatnya yang bertipe *intepreted*, menjadikannya bahasa yang ideal untuk *Scripting* dan pengembangan aplikasi secara cepat di banyak area pada banyak *platform*.

### 2.10.1 Nature Language ToolKit (NLTK)

*Nature Language Toolkit* (NLTK) adalah platform terkemuka untuk membangun program *Python* untuk bekerja dengan data mengenai bahasa manusia. NLTK menyediakan lebih dari 50 sumber daya *Corpora* dan *lexical* termasuk di dalamnya *WordNet*. Selain sumber daya tersebut NLTK juga dilengkapi dengan *library* yang biasa digunakan dalam pengolahan teks seperti *Classification*, *Tokenization*, *Stemming*, *Tagging*, *Parsing*, dan *Semantic Reasoning*, serta forum diskusi aktif yang membantu pengembangan dan

penggunaan NLTK (NLTK Project, 2017). Dalam penelitian ini NLTK dimanfaatkan untuk menyelesaikan tahap *Text Preprocessing* yakni tepatnya pada tahap Tokenisasi, lalu tahap *Stop-word Removal*.

### 2.10.2 Sastrawi

*Sastrawi* Python adalah *library* Python sederhana yang berfungsi untuk melakukan *Stemming*. *Stemming* adalah proses mengubah kata berimbuhan menjadi kata dasar. *Sastrawi* menggunakan kamus kata dasar dari kateglo.com dengan sedikit perubahan. (Librian, 2017). Pada penelitian ini *Sastrawi* digunakan dalam tahap *Text Preprocessing* tepatnya pada tahap *Stemming*.

### 2.10.3 Scikit-learn

Pedregosa et al (2011) membuat *Scikit-learn* sebagai sebuah modul Python yang mengintegrasikan berbagai algoritme *Machine Learning* untuk masalah *Supervised* dan *Unsupervised* dalam skala menengah. Modul ini berfokus pada membawa *Machine Learning* ke non-spesialis dengan menggunakan bahasa tingkat tinggi. *Scikit-learn* berhasil memiliki implementasi berbagai algoritme *Machine Learning*, baik *Supervised* maupun *Unsupervised*, menggunakan antarmuka berorientasi tugas yang konsisten, sehingga memungkinkan perbandingan metode yang mudah saat penggunaan. *Scikit-learn* juga mudah diintegrasikan dengan aplikasi di luar dari lingkup analisis data statistik dikarenakan diimplementasikan pada lingkungan Python. Pada penelitian ini *Scikit-learn* digunakan dalam proses *Term Weighting* dengan TF-IDF, dan klasifikasi sentimen yakni pada tahap pemilihan dan evaluasi model menggunakan *Kfold Cross-Validation*, dan klasifikasi menggunakan algoritme *Support Vector Machine*.

## 2.11 Visualisasi

Chen et al. (2007) mendefinisikan bahwa visualisasi data bukanlah sekedar merepresentasikan data dalam bentuk grafis. Lebih dari itu, informasi yang terkandung di dalamnya harus ditampilkan dalam tampilan yang baik, dalam artian dapat mengakomodasi pembaca agar dapat dengan mudah memahami struktur yang ada dalam data. Visualisasi data berkaitan dengan bidang visualisasi informasi (*Information Visualization*). Frasa informasi dalam visualisasi informasi menunjukkan bahwa visualisasi digunakan untuk semua jenis informasi, bukan hanya sekedar data.

### 2.11.1 Dashboard

Few (2006) mendefinisikan *Dashboard* sebagai tampilan visual dari informasi penting yang dibutuhkan dalam mencapai satu atau lebih tujuan. Informasi diatur sedemikian rupa sehingga dapat tampil dalam satu layar, dengan tujuan agar dapat dilihat dengan mudah. Malik (2005) menyatakan bahwa *Dashboard* dimanfaatkan untuk memonitor sistem yang kompleks dan *interdependent*. Pihak manajemen suatu organisasi memerlukan informasi yang

dibutuhkan guna menentukan kebijakan dan strategi untuk mendukung pencapaian tujuan organisasi. Sedangkan Eckerson (2005) menyatakan bahwa *Dashboard* didesain guna pencapaian tujuan strategis suatu organisasi atau perusahaan. Sehingga dengan penggunaan *Dashboard* diharapkan dapat mengukur, memonitor, dan memprediksi suatu aktivitas baik yang telah berjalan, sedang berjalan, maupun aktivitas selanjutnya. Oleh karena itu, organisasi diharapkan dapat mengukur dan memperbaiki strategi yang ada dalam rangka memaksimalkan kinerjanya.

Eckerson (2005) juga menyatakan kelebihan *Dashboard* di antaranya:

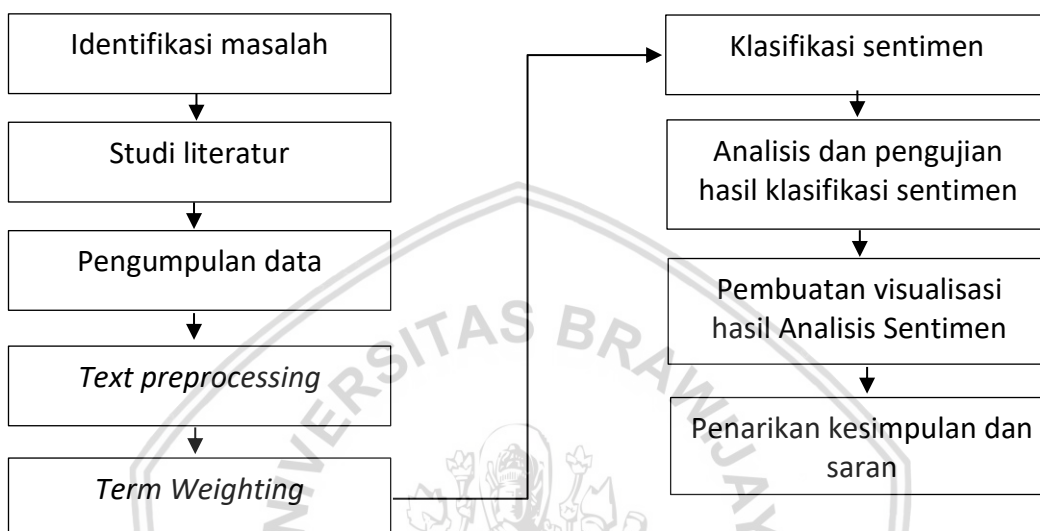
1. Alat untuk mengkomunikasikan strategi. Menjadi sarana untuk menyalurkan strategi yang dibuat oleh pihak eksekutif pada seluruh pihak yang memiliki peran.
2. Memonitor dan menyusun kembali pelaksanaan strategi melalui *monitoring* terhadap pelaksanaan strategi, pihak eksekutif dapat mengidentifikasi masalah kritis dan menentukan strategi sebagai solusinya.
3. Menyampaikan informasi kepada semua pihak sebagai penyaji informasi dalam satu layar yang menggunakan bagan, grafik, simbol, dan warna yang interaktif guna mempermudah pengguna dalam memahami persepsi informasi secara tepat.



## BAB 3 METODOLOGI

### 3.1 Metodologi penelitian

Alur dari penelitian Analisis Sentimen opini pelanggan terhadap aspek pariwisata pantai Malang Selatan menggunakan TF-IDF dan *Support Vector Machine* adalah sebagai berikut:



Gambar 3.1 Metodologi Penelitian

#### 3.1.1 Identifikasi masalah

Mengidentifikasi permasalahan terkait objek penelitian yakni pada pariwisata pantai Malang selatan. Masalah tersebut terkait dengan perlunya melakukan Analisis Sentimen opini pelanggan terhadap pariwisata pantai Malang Selatan untuk mendapatkan perspektif pelanggan.

#### 3.1.2 Studi literatur

Adanya studi literatur adalah untuk mencari teori – teori yang relevan terhadap penelitian sejenis yang pernah dilakukan sebelumnya yang berkaitan dengan penemuan sebelumnya, karena penelitian bertujuan untuk mendapatkan solusi dari permasalahan yang sedang dihadapi. Beberapa teori-teori yang digunakan seperti teori mengenai Opini, Analisis Sentimen, *Data Mining*, *Text Mining*, *Term Weighting*, *Machine Learning*, teori tentang pemilihan dan evaluasi model, serta visualisasi *Dashboard*.

#### 3.1.3 Pengumpulan data

Tahap ini meliputi pengumpulan data, pemilihan aspek, dan perancangan dataset. Pengumpulan data dilakukan terhadap data opini dalam bentuk ulasan yang diberikan pelanggan terhadap objek wisata pantai Malang Selatan. Data

bersumber dari situs TripAdvisor dengan menggunakan teknik *Web Scraping* untuk mendapatkan teks ulasan, pemberi ulasan serta waktu ulasan tersebut diberikan. *Tools* yang digunakan dalam proses ini adalah WebHarvy. Pemilihan aspek dilakukan dengan melakukan observasi terhadap data hasil proses *Web Scraping*, kemudian melakukan konfirmasi dengan Disparbud Kabupaten Malang. Perancangan *dataset* meliputi tahap pelabelan manual terhadap data opini. *Dataset* dilabelkan ke dalam tiga kelas sentimen yakni positif, negatif, dan netral. Positif untuk ulasan dengan sentimen positif terhadap aspek, negatif untuk ulasan dengan sentimen negatif terhadap aspek, netral untuk ulasan yang tidak memiliki sentimen terhadap aspek.

#### 3.1.4 Text Preprocessing

Sebelum dilakukan pengolahan, data terlebih dahulu melewati tahap *Text Preprocessing* untuk mengkondisikan data agar sesuai dengan kebutuhan untuk tahapan berikutnya yakni *Term Weighting* dan klasifikasi sentimen. Tahap *Text Preprocessing* meliputi pembuatan indeks kata (*Term indeks*), *Case Folding* untuk menyeragamkan huruf ke dalam bentuk *lowercase*, *Cleansing* untuk menghilangkan tanda baca yang tidak diperlukan, Tokenisasi untuk mengubah struktur teks ke dalam bentuk token. Formalisasi dan Translasi, untuk mengembalikan kata ke bentuk baku serta mengubah teks bahasa asing ke dalam bahasa Indonesia. *Stemming* untuk mengubah kata menjadi kata dasar, serta *Stop-word Removal* untuk menghilangkan kata yang masuk ke dalam kategori *stop-word*. Tahap *Text Preprocessing* menggunakan dalam Bahasa Pemrograman Python dengan memanfaatkan *module* yang sesuai seperti *NLTK*, *Sastrawi*, *Pandas*, dan lainnya.

#### 3.1.5 Term Weighting

*Term Weighting* atau pembobotan dilakukan dengan menerapkan *Term Frequency-Inverse Document Frequency* (TF-IDF). Pembobotan dilakukan dengan memanfaatkan *module* dalam Bahasa Pemrograman Python dengan *module Scikit-learn* untuk melakukan pembobotan TF-IDF.

#### 3.1.6 Klasifikasi sentimen

Berdasarkan hasil observasi data dan wawancara dengan dinas terkait, aspek yang dimiliki oleh pariwisata pantai Malang Selatan ada 5 aspek. Kelima aspek tersebut yakni akses Umum yang meliputi kondisi objek pariwisata pantai secara keseluruhan, aspek Kebersihan yang merupakan kondisi kebersihan dari objek pariwisata pantai, aspek Keramaian yang menunjukkan tingkat keramaian dari objek pariwisata pantai, aspek Akses Jalan yang menjelaskan kondisi akses jalan menuju objek pariwisata pantai, dan aspek kondisi Ombak yang menjelaskan kondisi ombak yang dimiliki objek pariwisata pantai. Pengujian model dan pemilihan *split data training* dan *testing* dengan menggunakan *Stratified K-Fold Cross Validation* dengan 4 *Fold* dan menghasilkan 75% data latih dan 25% data uji. Klasifikasi sentimen dilakukan dengan menggunakan algoritme *Support Vector*

*Machine* dengan menggunakan *Kernel Linear* dan strategi *Multiclass One-Vs-Rest* untuk menentukan sentimen dari opini terhadap tiap-tiap aspek yang telah ditentukan. Penerapan SVM dan *Stratified KFold Cross Validation* dilakukan dengan memanfaatkan *module Scikit-learn* pada Bahasa Pemrograman Python dengan menggunakan *LinearSVC* dan *StratifiedKFold*.

### 3.1.7 Analisis dan pengujian hasil klasifikasi sentimen

Selanjutnya dilakukan analisis hasil klasifikasi dan pengujian terhadap hasil dari klasifikasi sentimen pada tahap sebelumnya untuk mendapatkan nilai akurasi dari penggunaan algoritme *Support Vector Machine* yang dikombinasikan dengan *Term Frequency-Inverse Document Frequency*. Analisis yang dilakukan terkait hasil klasifikasi sentimen, perubahan sentimen dari waktu ke waktu, dan kontribusi objek pariwisata terhadap sentimen pada aspek. Jenis pengujian yang dilakukan adalah pengujian *F1-Score*, *Precision*, dan *Recall* serta penghitungan *Accuracy*. Pengujian dilakukan dengan memanfaatkan *module Scikit-learn* pada Bahasa Pemrograman Python.

### 3.1.8 Pembuatan visualisasi hasil Analisis Sentimen

Selanjutnya dilakukan visualisasi *Dashboard* untuk hasil Analisis Sentimen dan pengujian *Usability Dashboard* dengan menggunakan *System Usability Scale*. Visualisasi hasil Analisis Sentimen terdiri dari 4 *Dashboard* yakni *Dashboard* Analisis Sentimen menampilkan hasil klasifikasi sentimen opini pelanggan terhadap pariwisata pantai Malang Selatan. *Dashboard* trend sentimen memvisualisasikan perubahan sentimen opini pelanggan terhadap aspek pariwisata pantai Malang Selatan dari tahun 2013 hingga tahun 2018. *Dashboard* objek wisata adalah *Dashboard* yang menampilkan pemeringkatan objek pariwisata pantai Malang Selatan terhadap 5 aspek yang ada. *Dashboard* aspek memvisualisasikan data sentimen dari tahun 2013 hingga 2018 dari setiap objek wisata pantai Malang Selatan untuk setiap aspek.

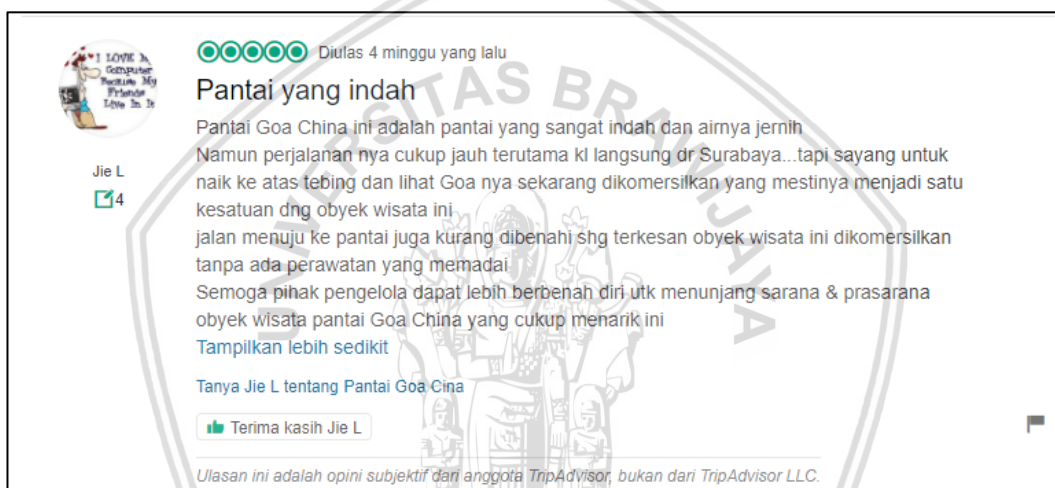
### 3.1.9 Penarikan kesimpulan dan saran

Tahap terakhir yaitu adalah penarikan kesimpulan berdasarkan rumusan masalah terkait Analisis Sentimen opini pelanggan terhadap aspek pariwisata pantai Malang Selatan dengan menggunakan TF-IDF dan *Support Vector Machine*. Kemudian pemberian saran untuk penelitian di masa mendatang

## BAB 4 PERANCANGAN

### 4.1 Pengumpulan data

Pengumpulan atau akuisisi data dilakukan dengan menggunakan metode *Web Scraping*. *Tool* yang digunakan dalam melakukan *Web Scraping* adalah WebHarvy. Data opini pelanggan pariwisata pantai Malang Selatan bersumber dari situs TripAdvisor. TripAdvisor adalah situs wisata yang memiliki 540 juta data ulasan dari 7,3 juta produk pariwisata termasuk didalamnya akomodasi, maskapai penerbangan, objek pariwisata, dan restoran (TripAdvisor, 2017). Gambar 4.1 memberikan contoh ulasan serta struktur ulasan dalam situs TripAdvisor berupa identitas pemberi ulasan, waktu pemberian ulasan, judul ulasan, *rating*, dan teks lengkap ulasan.

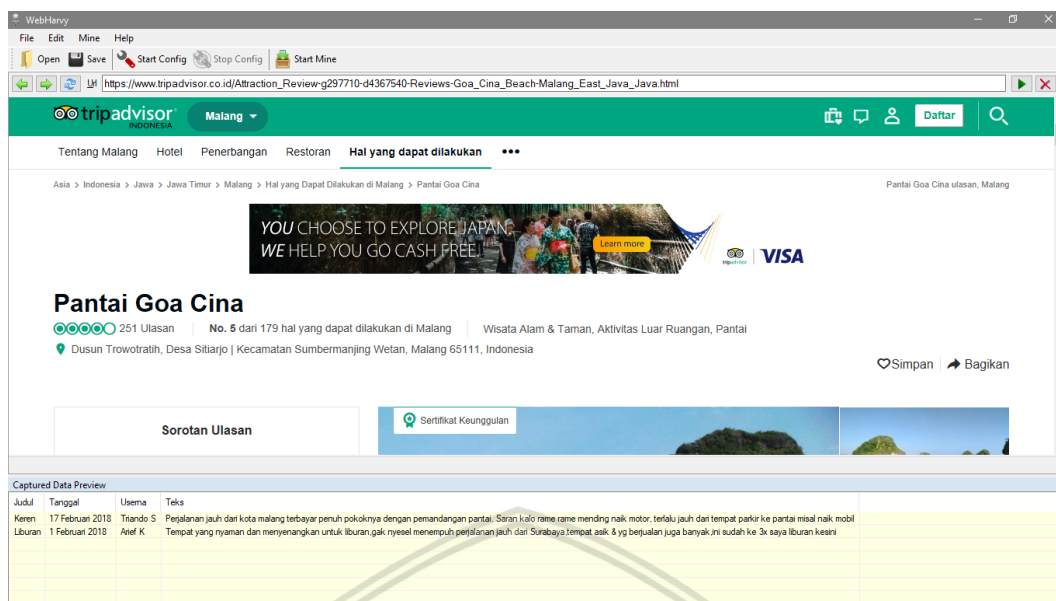


Gambar 4.1 Ulasan pada situs TripAdvisor

Proses *Scraping* yang dilakukan bertujuan untuk melakukan ekstraksi terhadap variabel yang dibutuhkan dalam proses Analisis Sentimen. Variabel tersebut antara lain adalah nama pemberi ulasan, waktu pemberian ulasan, judul, dan teks lengkap ulasan serta objek pariwisata yang diulas. Tahapan penerapan metode *Web Scraping* dengan menggunakan WebHarvy untuk data ulasan terhadap pariwisata pantai Malang Selatan yang bersumber dari situs TripAdvisor adalah sebagai berikut:

1. Membuka target *Uniform Resource Locator* (URL) situs yang berisi data yang akan diakuisisi

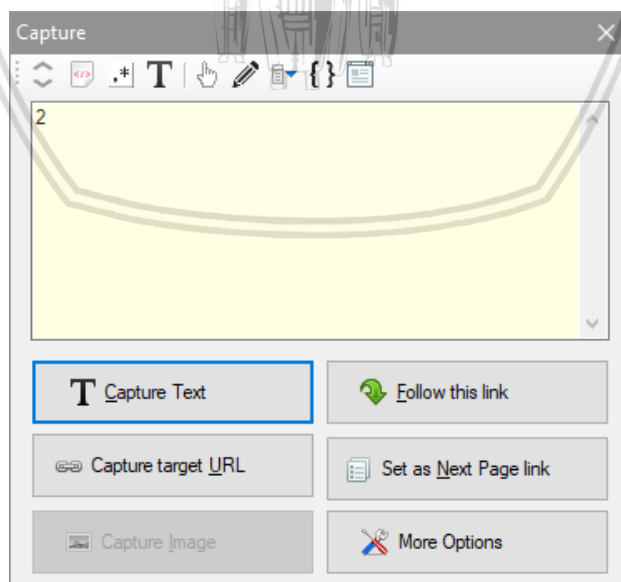
Untuk dapat melakukan *Scraping data* menggunakan WebHarvy yang perlu dipersiapkan adalah URL dari setiap objek pariwisata Pantai Malang Selatan. Gambar 4.2 menunjukkan contoh hasil URL target yang berhasil dibuka pada *Web browser* yang ada pada WebHarvy.



**Gambar 4.2 Load target URL pada WebHarvy**

2. Membuat konfigurasi untuk melakukan akuisisi data

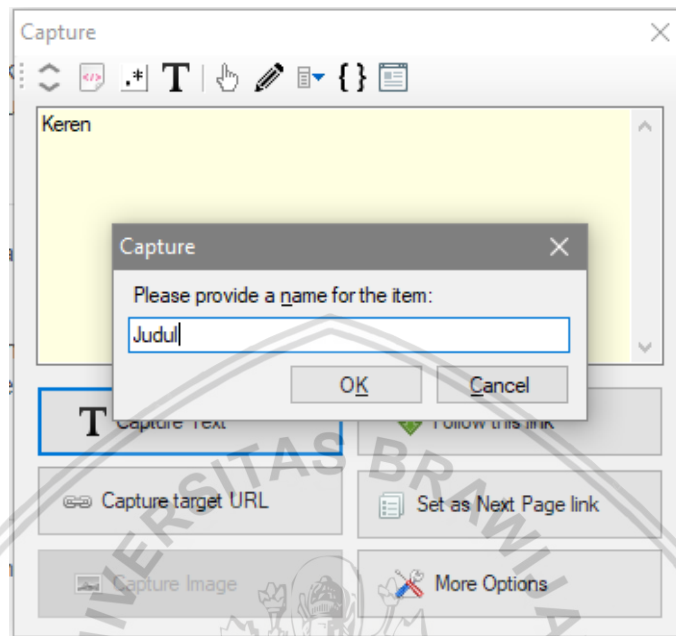
Untuk mendapatkan data dari URL target, perlu dilakukan konfigurasi terhadap variabel yang akan diakuisisi. Untuk memulai konfigurasi dengan mengaktifkan tombol “*Start Config*” pada bagian *toolbar*. Selanjutnya memulai konfigurasi dengan menerapkan tombol *next page* dengan tujuan mendapatkan data dari halaman-halaman berikutnya. Gambar 4.3 menunjukkan konfigurasi *next page* dengan melakukan “*Set as Next Page link*” untuk tombol 2.



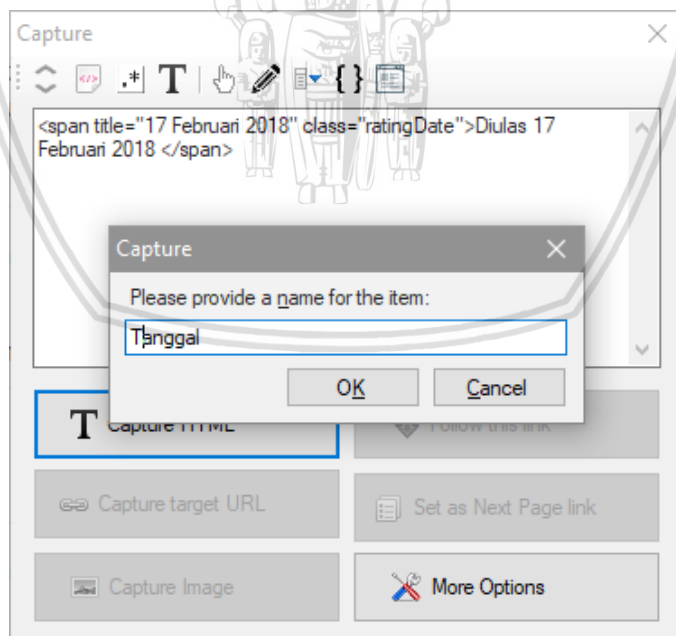
**Gambar 4.3 Konfigurasi next page**



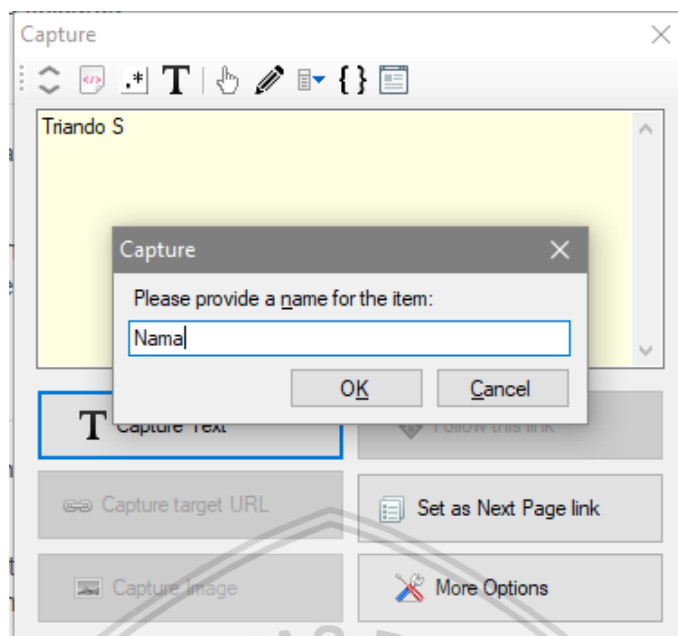
Tahapan selanjutnya ditunjukkan masing-masing pada gambar 4.4 untuk konfigurasi variabel Judul, Gambar 4.5 untuk konfigurasi variabel Tanggal, Gambar 4.6 untuk konfigurasi variabel Nama, Gambar 4.7 untuk variabel Teks.



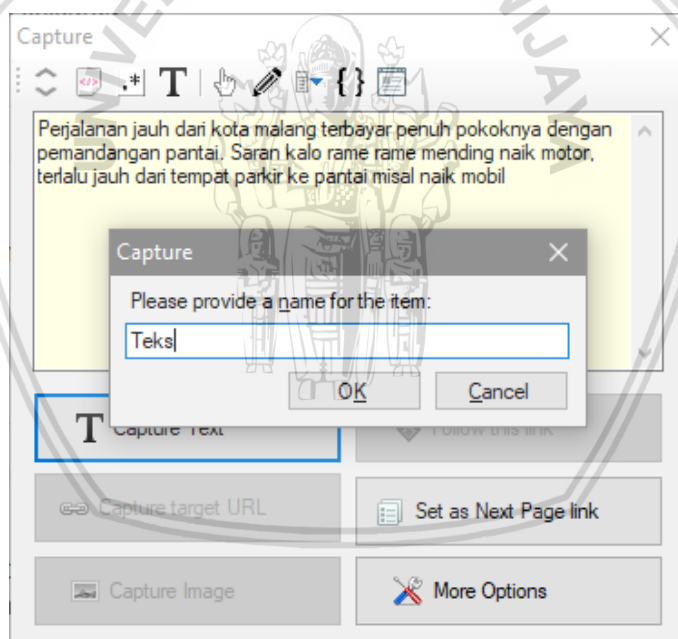
**Gambar 4.4 Konfigurasi *Scraping* variabel Judul**



**Gambar 4.5 Konfigurasi *Scraping* variabel Tanggal**



**Gambar 4.6 Konfigurasi *Scraping* variabel Nama**



**Gambar 4.7 Konfigurasi *Scraping* variabel Teks**

3. Memulai proses akuisisi data

Setelah selesai melakukan konfigurasi, WebHarvy akan menunjukkan *preview* data yang akan diakuisisi pada bagian bawah antarmuka WebHarvy. Gambar 4.8 menunjukkan hasil *preview* setelah konfigurasi sebelumnya disimpan.

Captured Data Preview			
Judul	Tanggal	Nama	Teks
Keren	17 Februari 2018	Triando S	Perjalanan jauh dari kota Malang terbayar penuh pokoknya dengan pemandangan pantai. Saran kalo rame rame mending naik motor, terlalu jauh dari tempat parkir ke pantai misal naik mobil
Liburan	1 Februari 2018	Anief K	Tempat yang nyaman dan menyenangkan untuk liburan, gak nyesel menempuh perjalanan jauh dari Surabaya, tempat asik & yg dijual juga banyak, ini sudah ke 3x saya liburan kesini

**Gambar 4.8 Preview data hasil konfigurasi *Scraping***

Setelah memastikan konfigurasi telah sesuai maka proses *Scraping* dapat dimulai dengan memilih *button* “*Start Mine*” pada bagian *toolbar* dari antarmuka WebHarvy. Kemudian WebHarvy akan menampilkan jendela yang memuat proses dan hasil *Scraping*. Pada jendela ini pula di definisikan mengenai jumlah halaman yang akan dilakukan proses *Scraping*, dan untuk memulai proses *Scraping* dengan menekan tombol “*Start*” Gambar 4.9 menunjukkan antarmuka jendela tersebut.



**Gambar 4.9 Antar muka jendela proses *Scraping* pada WebHarvy**

#### 4. Menyimpan hasil akuisisi data

Gambar 4.10 menunjukkan hasil yang muncul pada jendela proses *Scraping* setelah proses tersebut selesai. Kemudian data hasil proses *Scraping* kemudian disimpan dalam format csv dengan memilih tombol “*Export*” seperti pada Gambar 4.11. Sedangkan hasil akhir dari proses *Scraping* menghasilkan data ulasan variabel seperti pada Tabel 4.1 dengan menambahkan variabel Pantai secara manual pada file csv yang disimpan.

WebHarvy Miner

	Judul	Tanggal	Nama	Teks
1	Best place	16 September 2017	Ema S	Saat saya liburan saya bersama tema...
2	Mutiara tersembunyi	5 September 2017	Floren C	Sekarang udah banyak wisatawan y...
3	Liburan	4 September 2017	Amalia N	Waktu itu pengen banget kepantai . ...
4	Pantai Goa Cina ber-pasirputih dilinta...	19 Agustus 2017	Linna D	Sempet kesana dengan rombongan ...
5	mantai	21 Juli 2017	triasmidewi	letak nya jauh ke ujung mungkn kar...
6	Beautiful place to go	10 Juli 2017	bundaicha	Pantainya keren bingit...Ngga kalah ...
7	Pantainya bersih dan aimya jemih	9 Juli 2017	Leo Marta L	Keindahan pantainya dengan pasim...
8	Pantai yang bersih	6 Juli 2017	yulia r	Perjalanan 2 jam dari kota Malang. P...
9	Kecewa	17 Juni 2017	arinandah02	Pantai tetangga yang dikunjungi sete...
10	Eksotis	13 Mei 2017	Rikko A	Pantai goa cina terletak di malang sel...
11	goa cina malang	13 April 2017	afandi s	pengen ke pantai pasir putih tapi gak...
12	Pantai Gua Cina terlalu sederhana	11 April 2017	Kuntadi J	Pantai Gua Cina belum ada lima tahu...
13	Niceee	31 Maret 2017	Christina A	Medan kesana nya susahhhh...harus...
14	Honey moon II	17 Maret 2017	sayangabi	Ombak nya bagus banget... Besar... ..
15	Pantai Goa Cina	8 Maret 2017	DwiPuji25	Pantai yang bagus untuk menyaksik...
16	pantai selatan dengan ombak besar	5 Maret 2017	Ria L	2x kesana selalu ombak lagi besar, le...
17	Fantastic Beach	4 Maret 2017	Masduki68	Menyusuri destinasi wisata Kota Mala...
18	welcome goa cina malang	24 Februari 2017	windi a	lokasi very good, pemandangan v...
19	2 view yang cantik	21 Februari 2017	sylvianoerita	Goa China di selatan Malang ini meru...
20	Pemandangan indah	17 Februari 2017	Nicholas O	Pemandangan pantai yg sangat biru ...
21				

☒ Auto scroll above table ☐ Mine all pages  
 Number of pages to mine 2 Start Pause Stop Export

[20 Records mined from 2 Pages] Finished - Last Mined URL : [https://www.tripadvisor.co.id/Attraction\\_Review-g297710-d4367540-Reviews-or30-Goa\\_Cina...](https://www.tripadvisor.co.id/Attraction_Review-g297710-d4367540-Reviews-or30-Goa_Cina...)

Gambar 4.10 Tampilan jendela proses *Scraping* setelah proses selesai

Export Captured Data

Select File Format

CSV - Comma Separated Values

Save as

Export

Gambar 4.11 Penyimpanan data hasil *Scraping*

Tabel 4.1 Variabel dan tipe data hasil proses *Scraping*

Variabel	Tipe Data	Keterangan
Judul	String	Judul ulasan
Tanggal	Date	Waktu pemberian ulasan
Nama	String	Nama pemberi ulasan
Teks	String	Teks lengkap ulasan
Pantai	String	Objek pariwisata yang diulas

Tabel 4.2 merupakan contoh dari ulasan yang didapatkan dari proses *Scraping*. Proses *Scraping* menghasilkan 674 data ulasan berbahasa Indonesia yang diberikan oleh 423 pemberi ulasan pada 43 objek pariwisata pantai Malang Selatan. Data ulasan tersebut berjangka waktu dari tahun 2013 hingga tahun 2018.

**Tabel 4.2 Contoh data hasil proses *Scraping***

Judul	Tanggal	Nama	Teks	Pantai
ombak besar	19/06/2016	Tjutjukbs	Pantai ini terletak di pesisir selatan malang. Pantainya bersih dengan air yang cukup bersih juga. Pantai ini berombak cukup besar dan tidak disarankan untuk mandi ataupun berenang di pantai ini.	Ngantep
balekambang beach	14/10/2015	Ahmadaf rizal	serasa di kuta bali. pemandangan sunsetnya yang sangat indah membuat banyak wisatawan berbondong bondong mengunjungi pantai ini hanya sekedar menikmati indahnya matahari terbenam di pantai ini.	Balekambang
recommend	20/09/2016	Yulianile e	pantainya bagus tapi menurut saya jalan masuk ke pantainya sedikit kurang campur tangan pemerintah mohon di tingkatkan kembali :)	Gua Cina
Surga yg tersembunyi	30/01/2016	d2elysa	HTM 5rb biaya guide wajib 100rb untuk 10 orang. Nanti akan diantar dengan waktu trekking 1,5 jam mengunjungi pantai-pantai yg ada disana. Antara lain pantai tiga warna, watu pecah, gatra, dll.	Tiga Warna



## 4.2 Pemilihan aspek pariwisata pantai Malang Selatan

Berdasarkan observasi terhadap data ulasan yang ada di situs TripAdvisor serta setelah dikonfirmasi dengan proses wawancara dengan pihak Dinas Pariwisata dan Kebudayaan Kabupaten Malang Selatan didapatkan 5 aspek yang dimiliki oleh entitas pariwisata Pantai Malang Selatan. Tabel 4.3 menjelaskan mengenai nama serta deskripsi aspek yang dimiliki oleh pariwisata Pantai Malang Selatan.

**Tabel 4.3 Deskripsi aspek pariwisata pantai Malang Selatan**

Nama aspek	Deskripsi aspek
Umum	Menjelaskan kualitas secara umum yang dimiliki, termasuk di dalamnya adalah keindahan, kenyamanan, dari wisata pantai Malang Selatan, termasuk di dalamnya kelengkapan sarana prasarana pendukung.
Kebersihan	Menjelaskan kondisi tingkat kebersihan dari objek wisata pantai Malang Selatan.
Keramaian	Menjelaskan kondisi tingkat keramaian pengunjung dari objek wisata pantai Malang Selatan.
Akses jalan	Menjelaskan mengenai kondisi tingkat kemudahan akses menuju objek wisata pantai Malang Selatan.
Kondisi Ombak	Menjelaskan mengenai kondisi ombak dan perairan yang ada pada objek wisata pantai Malang Selatan.

## 4.3 Perancangan *dataset*

Tahap perancangan *dataset* meliputi tahap manualisasi pelabelan data yang dilakukan terhadap data teks uraian opini pelanggan untuk setiap aspek yang dimiliki oleh pariwisata pantai Malang Selatan. Label data dikelompokkan menjadi 3 kelas yang mewakili 2 sentimen yakni positif dan negatif terhadap aspek serta kelas netral yang menyatakan bahwa data tersebut tidak memiliki sentimen terhadap aspek. Untuk dapat menentukan label untuk setiap aspek disesuaikan dengan polaritas sentimen dari masing-masing aspek yang dijelaskan dalam Tabel 4.4 berikut ini.

**Tabel 4.4 Polaritas sentimen aspek pariwisata pantai Malang Selatan**

Nama aspek	Positif	Negatif
Umum	Menyatakan bahwa secara umum kualitas objek pariwisata sudah sesuai dengan ekspektasi atau keinginan pelanggan	Menyatakan bahwa secara keseluruhan kualitas objek pariwisata tidak dapat memenuhi ekspektasi atau keinginan pelanggan
Kebersihan	Menyatakan bahwa kondisi kebersihan maupun keasrian objek pariwisata pantai terjaga dengan baik	Menyatakan bahwa kondisi kebersihan tidak lagi terjaga dikarenakan adanya faktor-

		faktor yang merusak kebersihan
Keramaian	Menyatakan bahwa tingkat keramaian pengunjung masih rendah.	Menyatakan bahwa tingkat keramaian pengunjung tinggi
Akses jalan	Menyatakan bahwa kondisi sarana akses menuju lokasi pariwisata sudah baik dan memadai	Menyatakan bahwa kondisi sarana akses menuju lokasi pariwisata masih buruk dan menyulitkan
Kondisi Ombak	Menyatakan bahwa kondisi perairan dan ombak cukup tenang dan tidak berbahaya	Menyatakan bahwa kondisi perairan dan ombak membahayakan

Tabel 4.5 berikut ini merupakan contoh hasil pelabelan data untuk aspek Umum. Pelabelan ini di dasarkan pada polaritas sentimen yang sudah di definisikan sebelumnya. Untuk Kelas positif menyatakan bahwa sudah sesuai dengan ekspektasi pelanggan, dan untuk kelas negatif tidak sesuai dengan ekspektasi pelanggan. Untuk kelas netral berarti tidak memiliki ekspresi sentimen untuk aspek Umum.

**Tabel 4.5 Pelabelan data untuk aspek Umum**

Ulasan	Label
Pantainya luas dengan pasir putih yang bisa dijadikan tempat bermain pasir bahkan voli. Bisa juga untuk selancar (kemarin sih ada yang selancar). Di ujung timur terdapat aliran sungai yg berasal dari danau kecil di area pantai. Tersedia pula penginapan, camping ground, dan tempat ibadah.	Positif
Pantainya mempunyai ombak yang ga terlalu besar. Sehingga aman untuk bermain air di tepi pantai. Namun harus tetap waspada ketika sedang bermain. Penjual makanan juga gampang ditemui jadi kamu ga bakal kelaperan.	Positif
Kalo kita ke sana pas musim panas aja ..karena klo musim ujan air lautnya keruh ...jadi waktu snorkling ikannya ga kelihatan ...o iya jangan lupa bawa roti untuk makan ikan supaya naik ke atas....selamat berlibur	Negatif
Disini terdapat konservasi pohon mangrove, pasirnya putih, HTM 5rb ajah. Sayang waktu kesana pas surut jadi kurang begitu indah. Sedikit agak kotor mungkin karena terseret ombak.	Negatif
Pantai ini adalah pantai yang paling enak untuk digunakan memarkir kendaraan dan berbilas saat anda ingin ke bolu-bolu dan banyu anjlok. Anda bisa sewa perahu di sini untuk berkunjung ke tiga tempat itu. 350rb/perahu untuk maksimal 10 orang. Pantai di sini sebetulnya biasa saja tapi pantai di sekitarnya bagus. Tempat bilas yg saya gunakan cukup bersih. Ada banyak warung yg menyediakan menu seafood.	Netral

Tabel 4.6 berikut ini merupakan contoh hasil pelabelan data untuk aspek Kebersihan. Pelabelan ini di dasarkan pada polaritas sentimen yang sudah di definisikan sebelumnya. Untuk Kelas positif menyatakan bahwa kondisi kebersihan maupun keasrian objek pariwisata pantai terjaga dengan baik, dan untuk kelas negatif bahwa kondisi kebersihan tidak lagi terjaga dikarenakan adanya faktor-faktor yang merusak kebersihan. Untuk kelas netral berarti tidak memiliki ekspresi sentimen untuk aspek Kebersihan.

**Tabel 4.6 Pelabelan data untuk aspek Kebersihan**

Ulasan	Label
Pantai ini terletak di pesisir selatan malang. Pantainya bersih dengan air yang cukup bersih juga. Pantai ini berombak cukup besar dan tidak disarankan untuk mandi ataupun berenang di pantai ini.	Positif
Pantai ini masih sepi dari wisatawan sehingga cocok untuk refreshing. Wisatawan dapat menikmati hamparan pasir putih yang halus ditambah dengan deburan ombak biru serta tempat yang tenang. Pantai ini memiliki air yang biru jernih serta kondisi bebas sampah. Pemandangan di pantai ini cukup cantik dengan hiasan bunga pandan . Wisatawan dapat menyalurkan hobi surfing karena ombaknya yang besar sehingga cocok untuk tempat berselancar. Untuk yang hobi berenang kalau berenang disini berbahaya sebagai gantinya berenang di muara sungai yang berada di sebelah timur pantai Ngantep.	Positif
Sebenarnya asir lautnya cukup bersih tetapi pantainya agak kotor dengan sampah, sepertinya kurang dibersihkan. sayang sekali kurang diperhatikan	Negatif
Disini terdapat konservasi pohon mangrove, pasirnya putih, HTM 5rb ajah. Sayang waktu kesana pas surut jadi kurang begitu indah. Sedikit agak kotor mungkin karena terseret ombak.	Negatif
Pantainya mempunyai ombak yang ga terlalu besar. Sehingga aman untuk bermain air di tepi pantai. Namun harus tetap waspada ketika sedang bermain. Penjual makanan juga gampang ditemuin jadi kamu ga bakal kelaperan.	Netral

Tabel 4.7 berikut ini merupakan contoh hasil pelabelan data untuk aspek Keramaian. Pelabelan ini di dasarkan pada polaritas sentimen yang sudah di definisikan sebelumnya. Untuk Kelas positif menyatakan bahwa tingkat keramaian pengunjung masih rendah, dan untuk kelas negatif menyatakan bahwa tingkat keramaian pengunjung tinggi. Untuk kelas netral berarti tidak memiliki ekspresi sentimen untuk aspek Keramaian.

**Tabel 4.7 Pelabelan data untuk aspek Keramaian**

Ulasan	Label
Pantai ini masih sepi dari wisatawan sehingga cocok untuk refreshing. Wisatawan dapat menikmati hamparan pasir putih	Positif

yang halus ditambah dengan deburan ombak biru serta tempat yang tenang. Pantai ini memiliki air yang biru jernih serta kondisi bebas sampah. Pemandangan di pantai ini cukup cantik dengan hiasan bunga pandan . Wisatawan dapat menyalurkan hobi surfing karena ombaknya yang besar sehingga cocok untuk tempat berselancar. Untuk yang hobi berenang kalau berenang disini berbahaya sebagai gantinya berenang di muara sungai yang berada di sebelah timur pantai Ngantep.	
jalan menuju pantai ini adalah perempatan terakhir sebelum menuju ke balekambang anda ambil arah ke gua cina, kurang lebih 3 km akan ada petunjuk sebelah kanan pantai ngantep, pantai tidak terlalu banyak pengunjung namun cukup indah dengan air bening namun ombak agak besar, dipantai ini sudah dikelola ada tarif masuk 6500 per orang dan kendaraan roda 4 tarif 5000 rupiah	Positif
Untuk menjangkau lokasi ini harus menggunakan kendaraan pribadi. Mobil yg kuat untuk menjelajah medan tangguh sangat di sarankan. Karena mempunyai jalur yang berkelok-kelok dan mempunyai turunan yang cukup tajam. Untuk pantainya cukup ramai dikunjungi ketika pada waktu musim liburan. Toiletnya juga cukup memadai dan bersih. Overall wajib untuk dikunjungi.	Negatif
Pantai yang cukup jauh dari kota Malang dan membutuhkan kendaraan pribadi untuk menjangkaunya. Pantai ini bagi saya tidak terlalu menarik bagi saya karena cukup ramai sehingga kurang bisa menikmatinya. Akankah lebih baik jika anda ke sini saat hari biasa untuk menikmatinya lebih puas.	Negatif
dari pantai goa cina lurus dikit akan ketemu dengan pantai ini, pantai ungasan ... pantainya cukup luas dan pasirnya ok, sayang untuk tempat makan masih kurang .. tetapi ada pohon besar2 cocok bersantai sambil piknik ^ _ ^	Netral

Tabel 4.8 berikut ini merupakan contoh hasil pelabelan data untuk aspek Akses Jalan. Pelabelan ini di dasarkan pada polaritas sentimen yang sudah di definisikan sebelumnya. Untuk Kelas positif menyatakan bahwa kondisi sarana akses menuju lokasi pariwisata sudah baik dan memadai, dan untuk kelas negatif menyatakan bahwa kondisi sarana akses menuju lokasi pariwisata masih buruk dan menyulitkan. Untuk kelas netral berarti tidak memiliki ekspresi sentimen untuk aspek Akses Jalan.

**Tabel 4.8 Pelabelan data untuk aspek Akses Jalan**

Ulasan	Label
Pantai Ngantep terletak di Dusun Sukorejo, Desa Tumpakrejo, Kecamatan Gedangan, Kabupaten Malang, Jawa Timur. Pantai ini di kelola oleh pihak perhutani dan masyarakat sekitar Pantai Ngantep, jalan menuju Pantai Ngantep cukup mudah di lalui	Positif

dengan kendaraan pribadi. Menikmati hamparan pasir putih yang halus ditambah dengan deburan ombak biru serta tempat yang tenang akan menambah keinginan hati siapa saja untuk kembali lagi menyaksikan keindahan Pantai Ngantep.	
Pantai Ungapan di Malang Selatan, jarak dari kota Malang sekitar 80 km dapat ditempuh dengan mobil pribadi sekitar 2 jam. Akses jalan menuju Pantai Ungapan cukup bagus, Pantainya indah, terdapat muara yang dapat digunakan untuk bermain air atau berenang. Sekitar area pantai tersedia juga warung-warung makan dan area berkemah. Tersedia juga toilet/kamar mandi yang cukup bersih.	Positif
Destinasi wisata pantai yang berbeda di Malang selatan dengan panorama yg indah laut yg tenang diantara gugusan pulau pulau kecil sangat cocok dan aman untuk berenang recomanded banget untuk dikunjungi. Cuma perjalanan dr malang kota ke lokasi jauh banget, masih harus jalan kaki 1.5 km lagi. Tapi keren banget pemandangannya.	Negatif
Untuk menjangkau lokasi ini harus menggunakan kendaraan pribadi. Mobil yg kuat untuk menjelajah medan tangguh sangat di sarankan. Karena mempunyai jalur yang berkelok-kelok dan mempunyai turunan yang cukup tajam. Untuk pantainya cukup ramai dikunjungi ketika pada waktu musim liburan. Toiletnya juga cukup memadai dan bersih. Overall wajib untuk dikunjungi.	Negatif
Pantai ungapan ini cukup bagus untuk dikunjungi, pantainya cukup luas dan pasirnya juga bagus untuk bermain. untuk berkunjung ke pantai ini disarankan membawa perbekalan dari rumah, dikarenakan masih jarang .tempat makan di area pantai ini	Netral

Tabel 4.9 berikut ini merupakan contoh hasil pelabelan data untuk aspek kondisi Ombak. Pelabelan ini di dasarkan pada polaritas sentimen yang sudah di definisikan sebelumnya. Untuk Kelas positif menyatakan bahwa kondisi perairan dan ombak cukup tenang dan tidak berbahaya, dan untuk kelas negatif menyatakan bahwa kondisi perairan dan ombak membahayakan. Untuk kelas netral berarti tidak memiliki ekspresi sentimen untuk aspek Ombak.

**Tabel 4.9 Pelabelan data untuk aspek akses Ombak**

Ulasan	Label
Dipantai ini, wisatawan dapat menikmati sunset, atau berkemah. Kemah dibawa sendiri. Selain area pantai yang airnya relatif tenang, juga, ada 2 bukit karang yang membentang di tepi samudera indonesia. Dilihat dari kejauhan seperti Pintu Terbuka Lebar, itu sebab dinamai Clungup. Karang disini keren sekali, akibat diterpa air pantai. Meski airnya tak terlalu asin. Tiket masuk wisata ini, 1 paket dengan Pantai 3	Positif



Warna Rp.100.000,- / 1-10 orang dan asuransi Rp. 5.000,-/ per orang, dengan 1 guide. Biaya terpisah bila tidak mengunjungi pantai 3 warna. Aktivitas berenang ,bermain air, berkemah, menunggu sunset.	
Pantai Clungup ini sangat asri karena masih jarang dikunjungi wisatawan sehingga suasananya masih sangat bersih , nyaman dan indah. Hamparan pasir sangat luas bisa wisatawan jumpai disini. Pantai Clungup cukup tenang sampai jarak sekitar 500 m karena pantai dilindungi beberapa pulau kecil di lepas pantai , kedalamannya cukup dangkal sehingga bisa untuk berenang di sepanjang tepian pantai. Panorama pantai cukup indah apalagi pemandangan pulau kecil dan karang yang akan memanjakan mata wisatawan. Di sekitar pantai terdapat hutan mangrove yang madih lebat. Tempat ini bisa juga untuk camping. Selain keindahan pantainya suasananya yang sepi membuat wisatawan tensng. Nikmati juga sunset yang indah disini.	Positif
Pantai ini sekompleks dengan pantai batu bengkung. Setelah masuk gerbang masuk, pantai pertama yang terlihat adalah pantai ngopet. Pasir putih, dengan ombak yang besar. Pemandangannya sangat bagus!	Negatif
Pantai Ngopet bisa di akses dari dua jalur yang berbeda , pertama bisa melewati jalur arah Balekambang dan kedua dengan melewati jalur arah pantai Sendang Biru. Sebenarnya ombak disini tidak terlalu besar tetapi disarankan Wisatawan tidak berenang disini karena berhadapan langsung dengan laut lepas. Wisatawan disini disuguhi pemandangan eksotis batu koral dipadu dengan suasana bibir pantai yang masih hijau dan asri. Wisatawan bisa bermain pasir dan menikmati pemandangan pantai. Tempat ini juga bisa digunakan untuk camping.	Negatif
Pantai ini cukup indah letaknya bersebelahan dengan pantai gua cina dengan pulau karang di depannya dan suasana yang nyaman dan pasirnya halus, jika anda mau ke gua cina sekaligus mampir ke watu leter yang menjadi satu kawasan pantai	Netral

#### 4.4 Desain antar muka *Dashboard*

Hasil Analisis Sentimen divisualisasikan dalam 4 antar muka *Dashboard*, yakni *Dashboard* Analisis Sentimen opini terhadap aspek pariwisata Pantai Malang Selatan, *Dashboard trend* perubahan sentimen opini terhadap aspek pariwisata pantai Malang Selatan, *Dashboard* pemeringkatan objek pariwisata Pantai Malang Selatan terhadap 5 aspek yang ada, dan yang terakhir *Dashboard* frekuensi sentimen berdasarkan objek pariwisata per tahun untuk masing masing aspek.

Gambar 4.12 pada halaman 41 memperlihatkan desain antarmuka *Dashboard* hasil Analisis Sentimen opini pelanggan terhadap aspek pariwisata

Pantai Malang Selatan. Komponen *Dashboard* meliputi *bar chart* yang menampilkan hasil Analisis Sentimen terhadap semua aspek. Komponen legenda digunakan untuk melakukan interaksi terhadap data yang dimuat dalam *bar chart*. *Button* pada bagian atas dipergunakan untuk berpindah ke visualisasi *Dashboard* lain.

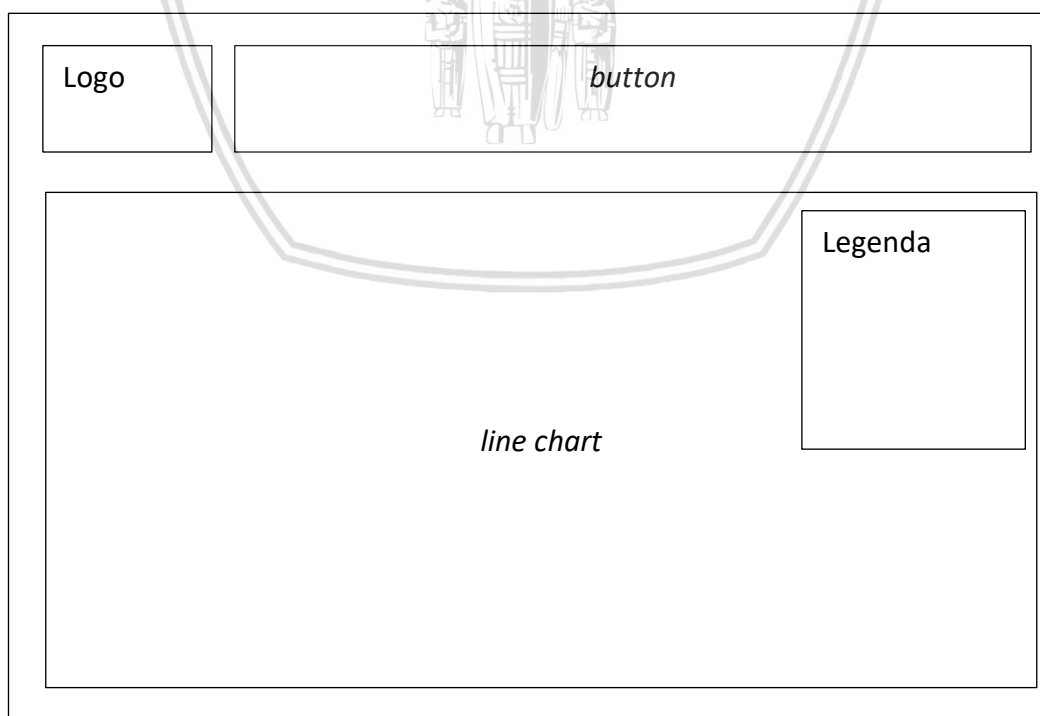
Gambar 4.13 pada halaman 41 memperlihatkan desain antarmuka *Dashboard* tren perubahan sentimen opini terhadap aspek pariwisata pantai Malang Selatan. Komponen *Dashboard* meliputi *line chart* yang menampilkan tren perubahan sentimen opini untuk setiap aspek dari tahun 2013 hingga tahun 2018. Komponen legenda digunakan untuk melakukan interaksi terhadap data yang dimuat dalam *line chart*. *Button* pada bagian atas dipergunakan untuk berpindah ke visualisasi *Dashboard* lain.

Gambar 4.14 pada halaman 42 memperlihatkan desain antarmuka *Dashboard* pemeringkatan objek pariwisata Pantai Malang Selatan terhadap 5 aspek yang ada. Komponen *Dashboard* meliputi *radar chart* yang nilai untuk setiap aspek yang dimiliki oleh objek wisata pantai Malang Selatan. Komponen legenda digunakan untuk melakukan interaksi terhadap data yang dimuat dalam *radar chart*. *Button* pada bagian atas dipergunakan untuk berpindah ke visualisasi *Dashboard* lain.

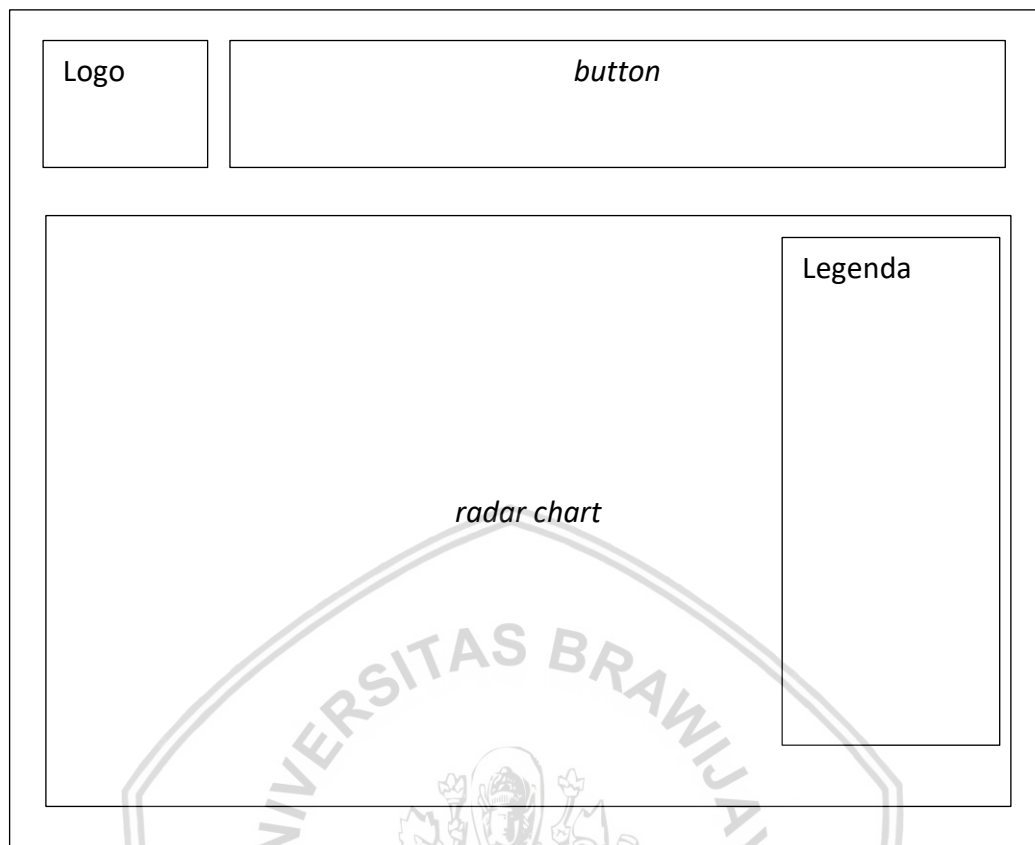
Gambar 4.15 pada halaman 42 memperlihatkan desain antarmuka *Dashboard* frekuensi sentimen berdasarkan objek pariwisata per tahun untuk masing masing aspek. Komponen *Dashboard* meliputi *bar chart* yang berisi data mengenai banyaknya frekuensi sentimen dari objek pariwisata terhadap masing-masing aspek dari tahun 2013 hingga 2018. *Button* pada bagian atas dipergunakan untuk berpindah ke visualisasi *Dashboard* lain.



**Gambar 4.12 Desain Dashboard Analisis Sentimen**



**Gambar 4.13 Desain antarmuka *Dashboard* tren perubahan sentimen**



**Gambar 4.14** Desain antarmuka *Dashboard* pemeringkatan objek wisata

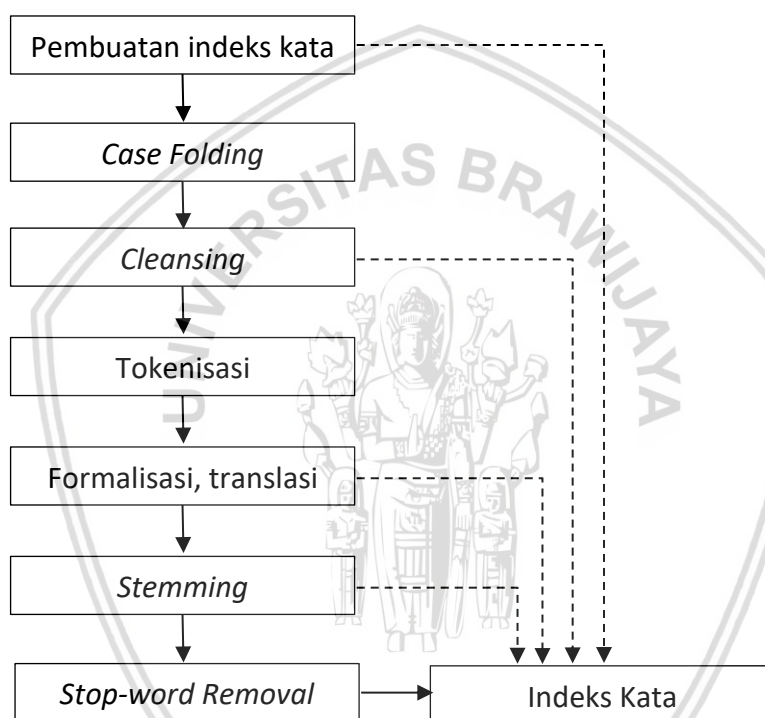


**Gambar 4.15** Desain antarmuka *Dashboard* pemeringkatan objek wisata

## BAB 5 IMPLEMENTASI

### 5.1 Text Preprocessing

Tahapan *Text Preprocessing* digunakan untuk mempersiapkan data sebelum diproses ke tahapan selanjutnya yakni *Term Weighting* dan klasifikasi sentimen. Tahapan ini menggunakan variabel teks ulasan sebagai teks yang akan diproses untuk kemudian menghasilkan indeks kata yang akan digunakan sebagai *features*. Tahapan *Text Preprocessing* dijelaskan oleh Gambar 5.1 meliputi pembuatan indeks kata, *Case folding*, *Cleansing*, tokenisasi, formalisasi dan translasi, *Stemming*, dan *Stop-word Removal*.



Gambar 5.1 Tahapan *Text preprocessing*

#### 5.1.1 Pembuatan indeks kata

Tujuan dari proses *Text Preprocessing* adalah mengubah teks dalam dokumen menjadi indeks kata yang mampu mewakili dokumen tersebut. Pembuatan indeks kata berjalan selama proses *Text Preprocessing*, hal tersebut merupakan upaya untuk membantu proses formalisasi yang dilakukan secara manual. Pembuatan indeks kata menggunakan cara menghitung kemunculan semua kata dalam semua dokumen ulasan. *Source code* 5.1 menjelaskan kode sumber yang digunakan untuk membuat indeks kata. Modul *counter* digunakan untuk menghitung kemunculan kata.

```

1 Import pandas as pd
2 From collections import Counter

```



```

3 Ulasan=pd.read_excel('ulasan.xlsx')
4 Termcount=[]
5 for j in ulasan['text']:
6     for k in j:
7         listcount.append(k)
8 counter=Counter(listcount)
9 final df = pd.DataFrame.from_dict(counter, orient='index')

```

### Source code 5.1 Implementasi pembuatan indeks kata (*Term indexing*)

Proses pembuatan indeks kata sebelum melakukan *Text Preprocessing* menghasilkan 4000 lebih variasi kata. Setelah melewati tahap *cleansing* serta formalisasi dan translasi jumlah variasi kata dalam indeks kata menjadi 3233 variasi kata. Tahap *stemming* mampu mereduksi variasi kata menjadi 1653 variasi kata. Terakhir variasi kata yang dihasilkan setelah proses *stop-word removal* sejumlah 1639 variasi kata.

#### 5.1.2 Case folding

*Case folding* bertujuan untuk mengubah seluruh huruf menjadi bentuk huruf kecil (*lowercase*). Implementasi dari proses *Case folding* ditunjukkan dalam *Source code* 5.2. Sedangkan Tabel 5.1 memberikan gambaran mengenai hasil dari implementasi *Case folding* yang diterapkan terhadap data teks lengkap ulasan.

```

1 Import pandas as pd
2 Df=pd.read_csv('ulasan.xlsx')
3 Df['Review']=df['Review'].str.lower()

```

### Source code 5.2 Implementasi Case folding

Tabel 5.1 Hasil implementasi *Case folding*

Sebelum <i>Case folding</i>	Sesudah <i>Case folding</i>
Pantai ini terletak di pesisir selatan malang. Pantainya bersih dengan air yang cukup bersih juga. Pantai ini berombak cukup besar dan tidak disarankan untuk mandi ataupun berenang di pantai ini.	pantai ini terletak di pesisir selatan malang. pantainya bersih dengan air yang cukup bersih juga. pantai ini berombak cukup besar dan tidak disarankan untuk mandi ataupun berenang di pantai ini.
serasa di kuta bali. pemandangan sunsetnya yang sangat indah membuat banyak wisatawan berbondong bondong mengunjungi pantai ini hanya sekedar menikmati indahnya matahari terbenam di pantai ini.	serasa di kuta bali. pemandangan sunsetnya yang sangat indah membuat banyak wisatawan berbondong bondong mengunjungi pantai ini hanya sekedar menikmati indahnya matahari terbenam di pantai ini.
pantainya bagus tapi menurut saya jalan masuk ke pantainya sedikit kurang campur tanggan pemerintah mohon di tingkatkan kembali :)	pantainya bagus tapi menurut saya jalan masuk ke pantainya sedikit kurang campur tanggan pemerintah mohon di tingkatkan kembali :)

HTM 5rb biaya guide wajib 100rb untuk 10 orang. Nanti akan diantar dengan waktu trekking 1,5 jam mengunjungi pantai-pantai yg ada disana. Antara lain pantai tiga warna, watu pecah, gatra, dll.	htm 5rb biaya guide wajib 100rb untuk 10 orang. nanti akan diantar dengan waktu trekking 1,5 jam mengunjungi pantai-pantai yg ada disana. antara lain pantai tiga warna, watu pecah, gatra, dll.
--	--

### 5.1.3 Cleansing

*Cleansing* merupakan proses yang bertujuan untuk menghilangkan elemen dalam teks yang dianggap sebagai *noise*. Pada praktiknya elemen seperti *HTML tag*, *url*, tanda baca, dan angka yang ada dalam teks dianggap sebagai *noise* sehingga akan dihilangkan. Implementasi dari proses *Cleansing* ditunjukkan dalam *Source code* 5.3. Sedangkan Tabel 5.2 memberikan gambaran mengenai hasil dari implementasi *cleansing* yang diterapkan terhadap data teks lengkap ulasan.

```

1 import string
2 import pandas as pd
3 import re
4 df=pd.read_csv('ulasan.xlsx')
5 clean=[]
6 for i in df['Review']:
7     clean.append(i)
8 newcl=[]
9 for c in clean:
10     for char in string.punctuation:
11         c=c.replace(char,'')
12         c = re.sub(r'\d+', '', c)
13     newcl.append(c)
14 df['sent']=newcl

```

**Source code 5.3 Implementasi Cleansing**

**Tabel 5.2 Hasil implementasi Cleansing**

Sebelum Cleansing	Sesudah Cleansing
pantai ini terletak di pesisir selatan malang. pantainya bersih dengan air yang cukup bersih juga. pantai ini berombak cukup besar dan tidak disarankan untuk mandi ataupun berenang di pantai ini.	pantai ini terletak di pesisir selatan malang pantainya bersih dengan air yang cukup bersih juga pantai ini berombak cukup besar dan tidak disarankan untuk mandi ataupun berenang di pantai ini
serasa di kuta bali. pemandangan sunsetnya yang sangat indah membuat banyak wisatawan berbondong bondong menngunjungi pantai ini hanya sekedar menikmati indahnya matahari terbenam di pantai ini.	serasa di kuta bali pemandangan sunsetnya yang sangat indah membuat banyak wisatawan berbondong bondong menngunjungi pantai ini hanya sekedar menikmati indahnya matahari terbenam di pantai ini
pantainya bagus tapi menurut saya jalan masuk ke pantainya sedikit	pantainya bagus tapi menurut saya jalan masuk ke pantainya sedikit

kurang campur tangan pemerintah mohon di tingkatkan kembali :)	kurang campur tangan pemerintah mohon di tingkatkan kembali
htm 5rb biaya guide wajib 100rb untuk 10 orang. nanti akan diantar dengan waktu trekking 1,5 jam mengunjungi pantai-pantai yg ada disana. antara lain pantai tiga warna, watu pecah, gatra, dll.	htm rb biaya guide wajib rb untuk orang nanti akan diantar dengan waktu trekking jam mengunjungi pantai pantai yg ada disana antara lain pantai tiga warna watu pecah gatra dll

#### 5.1.4 Tokenisasi

Tokenisasi dilakukan untuk memisahkan semua kata yang dipisahkan oleh spasi atau *whitespace* dalam teks ulasan. Modul *word\_tokenize* yang berasal dari *module* NLTK pada Python digunakan untuk melakukan tokenisasi. Implementasi dari proses tokenisasi ditunjukkan dalam *Source code* 5.4. Sedangkan Tabel 5.3 memberikan gambaran mengenai hasil dari implementasi tokenisasi yang diterapkan terhadap data teks lengkap ulasan.

```

1 from nltk.tokenize import word_tokenize
2 import pandas as pd
3 df=pd.read_csv('ulasan.xlsx')
4 token=[]
5 for i in df['Review']:
6     word=word_tokenize(i)
7     token.append(word)
8 df['Token Kata']=token

```

**Source code 5.4 Implementasi tokenisasi**

**Tabel 5.3 Hasil implementasi tokenisasi**

Sebelum tokenisasi	Sesudah tokenisasi
pantai ini terletak di pesisir selatan malang pantainya bersih dengan air yang cukup bersih juga pantai ini berombak cukup besar dan tidak disarankan untuk mandi ataupun berenang di pantai ini	['pantai', 'ini', 'terletak', 'di', 'pesisir', 'selatan', 'malang', 'pantainya', 'bersih', 'dengan', 'air', 'yang', 'cukup', 'bersih', 'juga', 'pantai', 'ini', 'berombak', 'cukup', 'besar', 'dan', 'tidak', 'disarankan', 'untuk', 'mandi', 'ataupun', 'berenang', 'di', 'pantai', 'ini']
serasa di kuta bali pemandangan sunsetnya yang sangat indah membuat banyak wisatawan berbondong bondong mengunjungi pantai ini hanya sekedar menikmati indahnya matahari terbenam di pantai ini	['serasa', 'di', 'kuta', 'bali', 'pemandangan', 'sunsetnya', 'yang', 'sangat', 'indah', 'membuat', 'banyak', 'wisatawan', 'berbondong', 'bondong', 'menunjungi', 'pantai', 'ini', 'hanya', 'sekedar', 'menikmati', 'indahnya', 'matahari', 'terbenam', 'di', 'pantai', 'ini']

pantainya bagus tapi menurut saya jalan masuk ke pantainya sedikit kurang campur tanggan pemerintah mohon di tingkatkan kembali	['pantainya', 'bagus', 'tapi', 'menurut', 'saya', 'jalan', 'masuk', 'ke', 'pantainya', 'sedikit', 'kurang', 'campur', 'tanggan', 'pemerintah', 'mohon', 'di', 'tingkatkan', 'kembali']
htm rb biaya guide wajib rb untuk orang nanti akan diantar dengan waktu trekking jam mengunjungi pantai pantai yg ada disana antara lain pantai tiga warna watu pecah gatra dll	['htm', 'rb', 'biaya', 'guide', 'wajib', 'rb', 'untuk', 'orang', 'nanti', 'akan', 'diantar', 'dengan', 'waktu', 'trekking', 'jam', 'mengunjungi', 'pantai', 'pantai', 'yg', 'ada', 'disana', 'antara', 'lain', 'pantai', 'tiga', 'warna', 'watu', 'pecah', 'gatra', 'dll']

### 5.1.5 Formalisasi dan translasi

Tabel 5.4 menunjukkan contoh hasil tahapan formalisasi. Formalisasi adalah proses mengembalikan bentuk kata ke dalam bentuk baku yang sesuai dengan KBBI. Proses ini dilakukan secara manual dengan menggunakan indeks kata yang telah dibuat dan membandingkannya dengan kamus Bahasa Indonesia. Tabel 5.5 menunjukkan contoh hasil tahapan formalisasi. Proses translasi bertujuan untuk melakukan penerjemahan bahasa dari bahasa asing ke bahasa Indonesia dengan tanpa menghilangkan makna yang dimaksud.

**Tabel 5.4 Formalisasi kata**

Kata dalam teks ulasan	Bentuk baku
aer, aernya, airannya, airnyapun	Air
ayok, ayookk, ayooo	Ayo
gak, ga, tdk.	Tidak
Adl, adlh,	Adalah
bangett, bangett, bangett	Sangat

**Tabel 5.5 Translasi kata**

Kata bahasa asing dalam teks ulasan	Dalam Bahasa Indonesia
amazing, ammaizing	Menakjubkan
battery, battrey	Baterai
beautiful, beautifull, beautifulx, beautifuly	Indah
Beach	Pantai
Booking	pesan

### 5.1.6 Stemming

*Stemming* bertujuan untuk mengubah kata ke dalam bentuk dasarnya atau kata dasarnya. Implementasi dari proses *Stemming* menggunakan *module* Sastrawi pada Python. Implementasi dari proses *Stemming* ditunjukkan dalam

Source code 5.5. Sedangkan Tabel 5.6 memberikan gambaran mengenai hasil dari implementasi *Stemming* yang diterapkan terhadap data teks lengkap ulasan.

```

1 from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
2 import pandas as pd
3 df=pd.read_csv('ulasan.xlsx')
4 factory = StemmerFactory()
5 stemmer = factory.create_stemmer()
6 sents=[]
7 for i in df['sent']:
8     sentence=i
9     output = stemmer.stem(sentence)
10    sents.append(output)
11 df['stems']=sents

```

### Source code 5.5 Implementasi *Stemming*

**Tabel 5.6 Implementasi *Stemming***

Sebelum <i>Stemming</i>	Sesudah <i>Stemming</i>
['pantai', 'ini', 'terletak', 'di', 'pesisir', 'selatan', 'malang', 'pantainya', 'bersih', 'dengan', 'air', 'yang', 'cukup', 'bersih', 'juga', 'pantai', 'ini', 'berombak', 'cukup', 'besar', 'dan', 'tidak', 'disarankan', 'untuk', 'mandi', 'ataupun', 'berenang', 'di', 'pantai', 'ini']	['pantai', 'letak', 'pesisir', 'selatan', 'malang', 'pantai', 'bersih', 'air', 'cukup', 'bersih', 'pantai', 'ombak', 'cukup', 'besar', 'tidak', 'saran', 'mandi', 'renang', 'pantai']
['serasa', 'di', 'kuta', 'bali', 'pemandangan', 'sunsetnya', 'yang', 'sangat', 'indah', 'membuat', 'banyak', 'wisatawan', 'berbondong', 'bondong', 'menunjungi', 'pantai', 'ini', 'hanya', 'sekedar', 'menikmati', 'indahya', 'matahari', 'terbenam', 'di', 'pantai', 'ini']	['serasa', 'kuta', 'bali', 'pandang', 'sunset', 'sangat', 'indah', 'buat', 'banyak', 'wisatawan', 'bondong', 'bondong', 'kunjung', 'pantai', 'hanya', 'kadar', 'nikmat', 'indah', 'sunset', 'pantai']
['pantainya', 'bagus', 'tapi', 'menurut', 'saya', 'jalan', 'masuk', 'ke', 'pantainya', 'sedikit', 'kurang', 'campur', 'tangan', 'pemerintah', 'mohon', 'di', 'tingkatkan', 'kembali']	['pantai', 'bagus', 'turut', 'jalan', 'masuk', 'pantai', 'sedikit', 'kurang', 'campur', 'tangan', 'perintah', 'mohon', 'tingkat']
['htm', 'rb', 'biaya', 'guide', 'wajib', 'rb', 'untuk', 'orang', 'nanti', 'akan', 'diantar', 'dengan', 'waktu', 'trekking', 'jam', 'mengunjungi', 'pantai', 'pantai', 'yg', 'ada', 'disana', 'antara', 'lain', 'pantai', 'tiga', 'warna', 'watu', 'pecah', 'gatra', 'dll']	['tiket', 'ribu', 'biaya', 'pandu', 'wajib', 'ribu', 'orang', 'akan', 'antar', 'waktu', 'jalan', 'jam', 'kunjung', 'pantai', 'pantai', 'ada', 'sana', 'pantai', 'tiga', 'warna', 'watu', 'pecah', 'gatra', 'lain']



### 5.1.7 Stop-word Removal

*Stop-word Removal* bisa disebut juga atau termasuk dalam proses *Filtering*, tujuannya adalah menghilangkan kata-kata yang masuk ke dalam kategori *Stop-word* seperti yang, aku, ke, di, dan lain sebagainya. Implementasi dari proses *Stop-word Removal* ditunjukkan dalam *Source code* 5.6. Sedangkan Tabel 5.7 memberikan gambaran mengenai hasil dari implementasi *Stop-word Removal* yang diterapkan terhadap data teks lengkap ulasan. Indeks kata yang masuk ke dalam kategori *Stop-word* didapatkan dari *module* NLTK yang ada pada Python.

```

1 from nltk.corpus import stopwords
2 import pandas as pd
3 df=pd.read_csv('ulasan.xlsx')
4 stop = set(stopwords.words('indonesian'))
5 words=[]
6 for i in df['stemmed']:
7     tokens=word_tokenize(i)
8     wordlist=[]
9     for word in tokens:
10         if word not in stop:
11             wordlist.append(word)
12     words.append(wordlist)

```

**Source code 5.6 Implementasi Stop-word Removal**

**Tabel 5.7 Implementasi Stop-word removal**

Sebelum <i>Stop-word Removal</i>	Sesudah <i>Stop word Removal</i>
['pantai', 'letak', 'pesisir', 'selatan', 'malang', 'pantai', 'bersih', 'air', 'cukup', 'bersih', 'pantai', 'ombak', 'cukup', 'besar', 'tidak', 'saran', 'mandi', 'renang', 'pantai']	['pantai', 'letak', 'pesisir', 'selatan', 'malang', 'pantai', 'bersih', 'air', 'bersih', 'pantai', 'ombak', 'saran', 'mandi', 'renang', 'pantai']
['serasa', 'kuta', 'bali', 'pandang', 'sunset', 'sangat', 'indah', 'buat', 'banyak', 'wisatawan', 'bondong', 'bondong', 'kunjung', 'pantai', 'hanya', 'kadar', 'nikmat', 'indah', 'sunset', 'pantai']	['serasa', 'kuta', 'bali', 'pandang', 'sunset', 'indah', 'wisatawan', 'bondong', 'bondong', 'kunjung', 'pantai', 'kadar', 'nikmat', 'indah', 'sunset', 'pantai']
['pantai', 'bagus', 'turut', 'jalan', 'masuk', 'pantai', 'sedikit', 'kurang', 'campur', 'tangan', 'perintah', 'mohon', 'tingkat']	['pantai', 'bagus', 'jalan', 'masuk', 'pantai', 'campur', 'tangan', 'perintah', 'mohon', 'tingkat']
['tiket', 'ribu', 'biaya', 'pandu', 'wajib', 'ribu', 'orang', 'akan', 'antar', 'waktu', 'jalan', 'jam', 'kunjung', 'pantai', 'pantai', 'ada', 'sana', 'pantai', 'tiga', 'warna', 'watu', 'pecah', 'gatra', 'lain']	['tiket', 'ribu', 'biaya', 'pandu', 'wajib', 'ribu', 'orang', 'jalan', 'jam', 'kunjung', 'pantai', 'pantai', 'pantai', 'warna', 'watu', 'pecah', 'gatra']

## 5.2 Term Weighting

Perhitungan pembobotan kata (*term*) dilakukan dengan menerapkan skema pembobotan TF-IDF. Bobot suatu kata akan meningkat sebanding dengan banyaknya kemunculan kata tersebut dalam satu dokumen. Selain itu, penghitungan IDF akan menghasilkan bobot yang sesuai dengan tingkat keunikan dari kata tersebut dalam indeks kata. Persamaan yang digunakan adalah persamaan (2.1) untuk menghitung  $idf_t$  dan (2.2) untuk menghitung bobot  $tf-idf_{t,d}$ . Implementasi dari proses *Term Weighting* ditunjukkan dalam *Source code 5.7*. Proses *Term Weighting* menggunakan *TfidfVectorizer* yang ada pada *module Scikit-learn* untuk Python.

```

1 import pandas as pd
2 from sklearn.feature_extraction.text import TfidfVectorizer
3 df=pd.read_excel(r"E:\\Preprocessed Fine 1.xlsx", usecols='H')
4 token={}
5 token=df['sent'].to_dict()
6 #tokenizing function
7 def tokenize(text):
8     word=word_tokenize(text)
9     return word
10 #vectorizer TF-IDF
11 tfidf=TfidfVectorizer(tokenizer=tokenize, lowercase=False)
12 tfs=tfidf.fit_transform(token.values())
13 #eksporting to file
14 feature_names = tfidf.get_feature_names()
15 dense = tfs.todense()
16 denselist=dense.tolist()
17 dftfidf=pd.DataFrame(denselist, columns=feature_names)
18 dftfidf.to_excel(r"E:\\tfidf.xlsx")

```

**Source code 5.7 Implementasi *Term Weighting* menggunakan TF-IDF**

### 5.2.1 Manualisasi

Manualisasi memberikan gambaran terhadap pembobotan dengan menggunakan TF-IDF. Pada Tabel 5.8 menunjukkan manualisasi hasil perhitungan *Term Frequency* (TF), *Document Frequency* (DF) dan dilanjutkan dengan penghitungan *Inverse Document Frequency* (IDF).

**Tabel 5.8 Menghitung TF, DF, dan IDF**

Term	TF					DF	IDF $\text{Log}(N/DF)$
	d1	d2	d3	d4	d5		
pantai	2	2	3	4	5	5	0
kemah	1	0	0	1	0	2	0,39794
indah	0	0	0	0	1	1	0,69897
main	1	2	0	0	0	2	0,39794
selancar	2	0	1	0	0	2	0,39794
malang	0	0	1	1	1	3	0,2218487
pasir	2	0	0	0	1	2	0,39794
timur	1	0	0	0	1	2	0,39794

nikmat	0	0	0	1	1	2	0,39794
--------	---	---	---	---	---	---	---------

Dilanjutkan dengan menghitung bobot TF-IDF yang merupakan hasil perkalian dari TF dengan IDF. Berikut adalah contoh perhitungan bobot TF-IDF untuk kata 'selancar' pada d1:

$$Tf-idf_{t,d} = TF * IDF = 2 * 0,397940 = 0,795880017$$

Tabel 5.9 ditunjukkan hasil perhitungan bobot TF-IDF untuk setiap *term* untuk setiap dokumen.

**Tabel 5.9 Menghitung bobot TF-IDF**

Term	TF-IDF				
	d1	d2	d3	d4	d5
Pantai	0	0	0	0	0
kemah	0,39794	0	0	0,39794	0
indah	0	0	0	0	0,69897
main	0,39794	0,79588	0	0	0
selancar	0,79588	0	0,39794	0	0
malang	0	0	0,221849	0,221849	0,221849
pasir	0,79588	0	0	0	0,39794
timur	0,39794	0	0	0	0,39794
nikmat	0	0	0	0,39794	0,39794

### 5.3 Klasifikasi sentimen

Analisis sentimen dilakukan terhadap 5 aspek pariwisata pantai Malang Selatan. Aspek tersebut antara lain aspek Umum, Kebersihan, Keramaian, Akses Jalan, dan Ombak. Aspek Umum menggambarkan penilaian secara keseluruhan terhadap objek pariwisata pantai Malang Selatan. Aspek Kebersihan merujuk pada tingkat Kebersihan dari objek pariwisata pantai Malang Selatan. Aspek Keramaian merujuk pada tingkat keramaian dari objek pariwisata pantai Malang Selatan. Aspek Akses Jalan merujuk pada kondisi dari akses jalan menuju objek pariwisata pantai Malang Selatan. Aspek Ombak merujuk pada kondisi ombak dan perairan dari objek pariwisata pantai Malang Selatan

Sebelum melakukan klasifikasi sentimen terhadap setiap aspek pariwisata pantai Malang Selatan dibutuhkan 2 data. Data pertama adalah data ulasan yang telah diubah menjadi data bobot yang sebelumnya telah dibuat dalam tahap *Text Preprocessing* dan *Term Weighting*. Data kedua adalah data yang berisi label *class* dari data ulasan yang mewakili sentimen untuk setiap aspek pariwisata pantai Malang Selatan. *Class* sentimen dibagi menjadi positif, negatif, dan netral. Positif mengartikan bahwa ulasan memiliki sentimen positif terhadap aspek tersebut. Aspek negatif mengartikan bahwa ulasan memiliki sentimen positif terhadap

aspek tersebut. Netral mengartikan bahwa ulasan tidak memiliki atau tidak memberikan sentimen terhadap aspek tersebut.

Klasifikasi sentimen menggunakan pendekatan *Supervised Learning*. *Supervised learning* dilakukan dengan mengimplementasikan *Support Vector Machine* dengan memanfaatkan *module Scikit-learn* pada Python. *Classifier* yang digunakan adalah *LinearSVC* dengan strategi *multi class one versus rest*. Proses klasifikasi sentimen dibagi menjadi 2 tahapan yakni pemilihan model menggunakan *K-fold cross validation* dan klasifikasi sentimen menggunakan *LinearSVC*.

### 5.3.1 Pemilihan model menggunakan *K-fold Cross Validation*

Pemilihan model dilakukan untuk mendapatkan model terbaik dengan akurasi tertinggi untuk setiap aspek. Pemilihan model dilakukan dengan menggunakan metode *K-fold Cross Validation*. *K-fold Cross Validation* adalah proses sistematis untuk melakukan prosedur penyusunan *dataset* menjadi data latih dan data uji secara berulang-ulang dengan tujuan untuk menghindari terjadinya variansi tunggal yang terbentuk saat melakukan hanya dengan satu set data latih dan data uji.

*Source code* 5.8 menunjukkan implementasi dari pemilihan model menggunakan *K-Fold Cross Validation*. *Stratified K-fold* digunakan untuk memastikan rasio untuk semua *Class* seimbang untuk semua *fold*. *Dataset* dibagi menjadi 4 *Fold* untuk mendapatkan rasio data latih dan data uji sebesar 75:25. Parameter *shuffle* bernilai *True* dengan tujuan untuk mendapatkan model dengan variasi lebih banyak. *Parameter tuning* juga dilakukan untuk mendapatkan kombinasi parameter yang menghasilkan rata-rata akurasi tertinggi, *Parameter Tuning* dilakukan terhadap parameter dari *random\_state* dalam proses *K-fold*. Setelah mendapatkan parameter terbaik maka dilakukan pemilihan model dengan menjalankan *cross\_val\_score* untuk mendapatkan kombinasi data latih dan uji dengan akurasi tertinggi. Setelah mendapatkan kombinasi data latih dan uji yang memiliki akurasi tertinggi maka data tersebut disimpan ke dalam 4 variabel. 4 variabel tersebut adalah *X\_train*, *y\_train*, *X\_test*, dan *y\_test*. *X\_train* untuk data bobot ulasan yang digunakan sebagai data latih. *y\_train* untuk data label yang digunakan sebagai data latih. *X\_test* untuk data ulasan yang digunakan sebagai data uji. *y\_test* untuk data label yang digunakan sebagai data uji. Semua tahapan tersebut kemudian dilakukan terhadap data untuk masing-masing aspek untuk mendapatkan model terbaik untuk masing-masing aspek.

```

1 import pandas as pd
2 from sklearn.svm import LinearSVC
3 from sklearn.model_selection import StratifiedKFold
4 from sklearn import metrics
5 from sklearn.model_selection import cross_val_score
6 from sklearn.metrics import classification_report
7 from sklearn.metrics import confusion_matrix
8 review = pd.read_excel(r"E://Preprocessed   Fine   1.xlsx",
   sheet_name="Sheet3")

```

```

9  tfidf = pd.read_excel(r"E://tfidf.xlsx", sheet_name="Sheet1")
10 X = np.array(tfidf[:].values.tolist())
11 y = np.array(review['Aspek'].values.tolist())
12 svm=LinearSVC(multi_class='ovr')
13 para=[]
14 for i in range(0, 100):
15     kf = StratifiedKFold(n_splits=4, shuffle=True, random_state=i)
16     scores = cross_val_score(svm, X, y, cv=kf,
17                             scoring='accuracy').mean()
18     para.append(scores)
19 max_value = max(para)
20 max_index_random = para.index(max_value)
21 split=[]
22 kf = StratifiedKFold(n_splits=4, shuffle=True,
23                     random_state=max_index_random)
24 scores = cross_val_score(svm, X, y, cv=kf, scoring='accuracy')
25 split=scores.tolist()
26 max_value = max(split)
27 max_index_set = split.index(max_value)
28 print (max_index_random, max_index_set)
29 i=0;
30 for train_index, test_index in kf.split(X, y):
31     if i==max_index_set:
32         print("TRAIN:", train_index, "TEST:", test_index)
33         X_train, X_test = X[train_index], X[test_index]
34         y_train, y_test = y[train_index], y[test_index]
35         break
36     else:
37         i=i+1

```

**Source code 5.8 Implementasi Stratified K-Fold Cross Validation**

### 5.3.2 Klasifikasi sentimen menggunakan *LinearSVC*

Implementasi klasifikasi sentimen menggunakan *LinearSVC* yang bertindak sebagai *classifier*. Klasifikasi sentimen dilakukan dengan melakukan *fit* atau *training* dengan memanfaatkan variabel *X\_train* dan *y\_train* yang didapatkan dari proses sebelumnya. Hasil proses *fit* selanjutnya digunakan untuk melakukan klasifikasi sentimen dengan menggunakan fungsi *predict* dan mengirimkan variabel *X\_test* yang bertindak sebagai data yang akan diklasifikasikan. Hasil dari penggunaan fungsi *predict* berupa *array* yang berisi hasil klasifikasi sentimen dari variabel *X\_test*. Implementasi dari proses klasifikasi sentimen menggunakan *LinearSVC* ditunjukkan pada *Source code 5.9*.

```

1  from sklearn.svm import LinearSVC
2  svm_model_linear = LinearSVC(multi_class='ovr').fit(X_train, y_train)
3  pred = svm_model_linear.predict(X_test)

```

**Source code 5.9 Implementasi klasifikasi sentimen menggunakan *LinearSVC***



## BAB 6 ANALISIS DAN PEMBAHASAN

Analisis dari hasil perancangan dan implementasi dibahas pada bab ini. Pembahasan tersebut terkait hasil klasifikasi dan pengujian dari hasil klasifikasi sentimen untuk masing-masing aspek, implementasi *Dashboard* dan pembahasan mengenai hasil Analisis Sentimen terhadap aspek pariwisata pantai Malang Selatan yang divisualisasikan menggunakan *Dashboard*.

### 6.1 Hasil klasifikasi

Pada subbab ini membahas mengenai hasil klasifikasi sentimen untuk masing-masing aspek.

#### 6.1.1 Aspek Umum

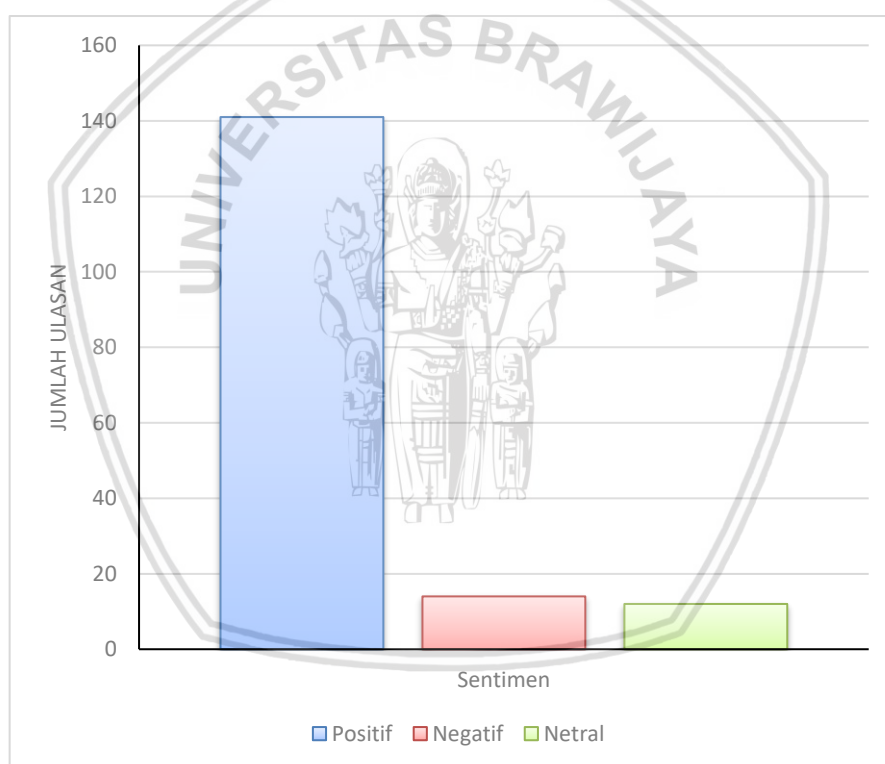
Aspek Umum meliputi kondisi keindahan dan kenyamanan dari wisata pantai Malang Selatan. Dalam aspek Umum, *class* positif berisi ulasan yang menyatakan tanggapan positif terhadap kondisi pantai secara keseluruhan. Sedangkan *class* negatif berisi ulasan yang menyatakan tanggapan negatif terhadap kondisi pantai secara keseluruhan. Tabel 6.1 menunjukkan contoh teks ulasan yang berhasil diklasifikasikan ke dalam 3 sentimen yang ada yakni positif negatif dan netral.

**Tabel 6.1 Hasil klasifikasi aspek Umum**

Teks	Kelas
Pantai tersebut tidak ada yang spesial kecuali ada gua kecil, pantainya juga banyak bebatuan, jadi kurang cocok untuk bersantai atau jalan-jalan. akses ke lokasi tersebut juga masih tidak baik, bisa merusak mobil. lebih baik pakai mobil sewaan ke tempat ini.	0 (Negatif)
Pantai yang cukup jauh dari kota Malang dan membutuhkan kendaraan pribadi untuk menjangkaunya. Pantai ini bagi saya tidak terlalu menarik bagi saya karena cukup ramai sehingga kurang bisa menikmatinya. Akankah lebih baik jika anda ke sini saat hari biasa untuk menikmatinya lebih puas.	0 (Negatif)
Pantai ini sekompleks dengan pantai batu bengkung. Setelah masuk gerbang masuk, pantai pertama yang terlihat adalah pantai ngopet. Pasir putih, dengan ombak yang besar. Pemandangannya sangat bagus!	1 (Positif)
Pantai mbehi juga dikenal dengan nama rante wulung, terletak dikawasan konservasi hutan lindung, untuk mencapai tempat ini harus ditemani oleh pemandu dan jalan naik turun melalui hutan sekitar 30 menit dari kondang sugu, indah sekali pemandangan disini dengan ombak yang besar, dan banyak kura kura	1 (Positif)

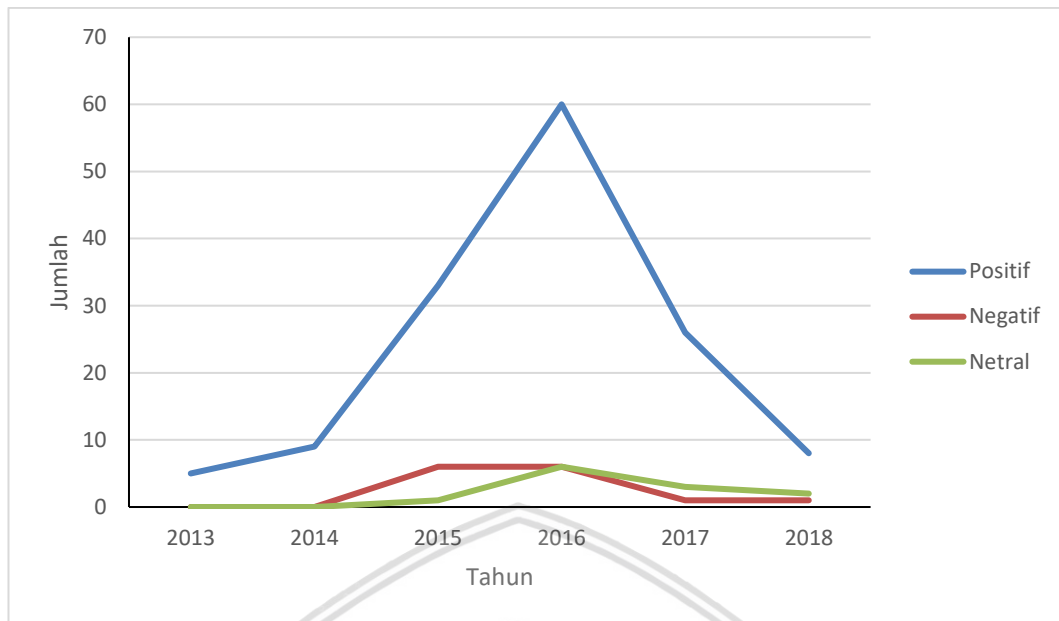
letaknya berdekatan dengan panatai kondang merak , ada beberapa penjual , jalanan lebih bagus dari pantai kondang merak. tidak disarankan mandi di laut jika ombak tinggi , ada pure yang terletak disebelah selatan yang mirip dengan Tanah Lot.	2 (Netral)
Salah satu wisata alternatif di Malang yang mulai ramai dikunjungi. Untuk berkunjung ke sini pengunjung dapat menggunakan perahu dari pantai Lenggoksono sekitar 20-30 menit. Dari jauh pengunjung sudah dapat melihat air terjun banyu anjlok yang airannya mengarah ke laut. Selain menikmati air terjun, pengunjung juga bisa menaiki tangga menuju ke atas untuk berenang di kolam alami yang terbentuk dari tumpahan air terjun	2 (Netral)

Pada Gambar 6.1 menunjukkan aspek Umum menghasilkan 14 ulasan dengan sentimen negatif, 141 ulasan dengan sentimen positif, dan 12 ulasan dengan sentimen netral.



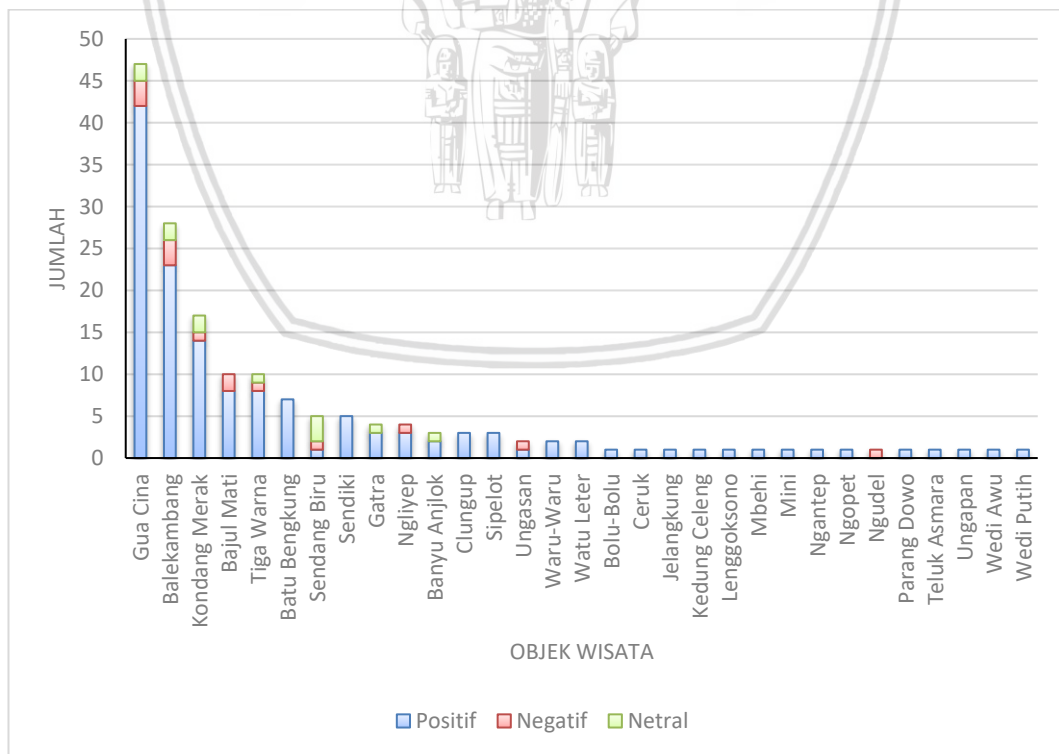
**Gambar 6.1 Frekuensi hasil klasifikasi terhadap aspek Umum**

Pada Gambar 6.2 memperlihatkan tren dari ulasan terhadap aspek Umum. Dari Gambar tersebut menunjukkan pada tahun 2013 hingga 2016 terjadi peningkatan ulasan positif terhadap aspek Umum dengan puncaknya hingga 60 ulasan pada tahun 2016. Sedangkan untuk tahun 2016-2018 ulasan positif cenderung turun hingga hanya 8 ulasan positif untuk tahun 2018. Ulasan negatif memuncak pada tahun 2015 dengan total 8 ulasan negatif namun mengalami penurunan hingga pada 2017 hanya ada 1 ulasan negatif, dan 2018 dengan 2 ulasan negatif.



**Gambar 6.2 Tren hasil klasifikasi terhadap aspek Umum**

Gambar 6.3 menunjukkan jumlah ulasan berdasarkan nama objek wisata pantai Malang Selatan. Dari hasil klasifikasi menunjukkan pantai Gua Cina, Balekambang, dan Kondang Merak menjadi 3 teratas pantai yang memiliki ulasan dalam aspek Umum.



**Gambar 6.3 Ulasan berdasarkan objek wisata terhadap aspek Umum**

### 6.1.2 Aspek Kebersihan

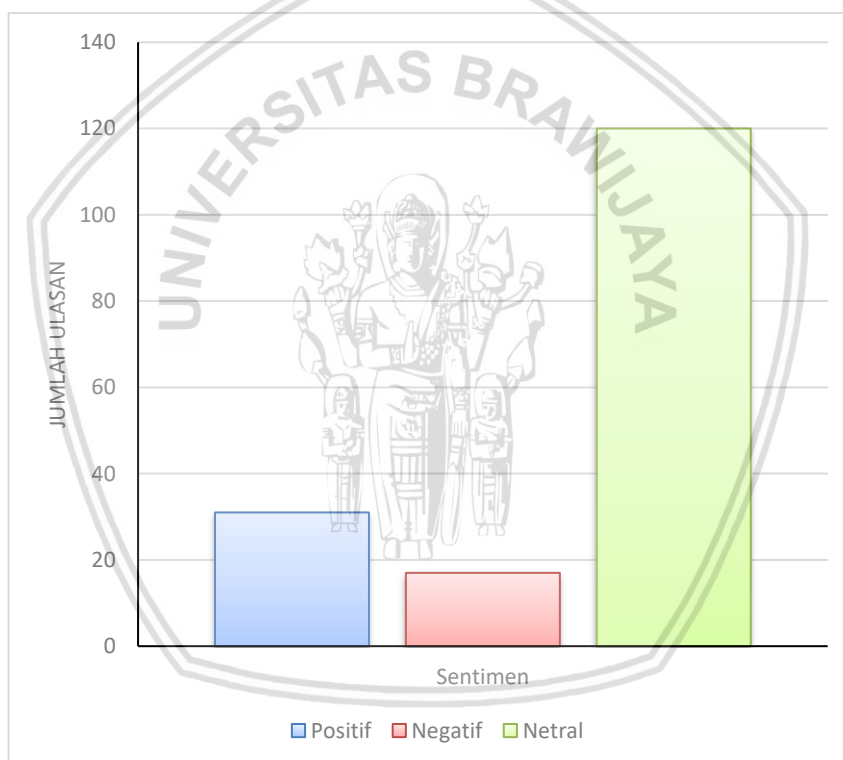
Aspek Kebersihan meliputi kondisi kebersihan dari objek wisata pantai Malang Selatan. Dalam aspek Kebersihan, *class* positif berisi ulasan yang menyatakan tanggapan positif terhadap kondisi kebersihan pantai. Sedangkan *class* negatif berisi ulasan yang menyatakan tanggapan negatif terhadap kondisi kebersihan pantai. Tabel 6.2 menunjukkan contoh teks ulasan yang berhasil diklasifikasikan ke dalam 3 sentimen yang ada yakni positif negatif dan netral.

**Tabel 6.2 Hasil klasifikasi aspek Kebersihan**

Teks	Kelas
semakin banyak orang meng share foto keindahan pantai nya, semakin banyak lagi pengunjung yang datang, semakin menggunung sampah2 nya. saya cuma prihatin,, ada dua sisi pantai di sepanjang balekambang, sebelah barat dekat dengan pura, sangat kotor karena pengunjung lebih banyak di situ belum lagi pedagang, hampir tidak ada tempat sampah, apalagi petugas kebersihan. Fasilitas toilet umum juga sepertinya dikelola para pemilik warung, dengan kebersihan ala kadarnya. sebelah timur pantai relatif lebih bersih, pemandangan lebih bagus, ada mangrove rafting juga tarifnya 25 ribu.	0 (Negatif)
Beautiful beach but rubbish every where, banyak sampah hampir di seluruh area .....banyak pungutan liar(bayar parkir 2 kali, di loket pintu masuk dan di area parkir di paksa bayar parkir lagi) pengelola harus banyak belajar tentang penataan dan pengelolaan wisata pantai, agar bisa menjadi seperti pantai2 di Bali, Lombok, Serta Pantai pulau merah Banyuwangi yg sudah mulai berkembang ..... karena Pantai2 di Malang tidak kalah bagus..... juga kesadaran para pengunjung untuk tidak membuang sampah sembarangan.....	0 (Negatif)
pantai goa cina menurut saya adalah salah satu pantai yang cukup indah, bersih , dan cantik .. menempuh perjalanan yang jauh dan melelahkan akan terbayarsetelah sampai di pantai ini .. airnya yang masih biru dan benar-benar bersih membuat pantai ini terlihat sangat menarik ,, sungguh ,enyenangkan melihat dan menikmati pemandangan yang ada dipantai ini ,, tidak akan menyesal berkujung ke pantai ini dan saya sangat ingin kembali lagi ke pantai goa cina	1 (Positif)
Pantai ini sangat bagus dan sangat mempesona karena masih alami dan pasirnya putih bersih ini perlu di pertahankan kebersihaannya, kalau di dibandingkan pantai-pantai yang lain yang ada di selatan kota malang ini hamya akses masuk kepantai perlu di perlebar jalannya	1 (Positif)
Pertama kali datang ke pantai ini sama teman kantor pada 2/2016. Ruar biasa senangnya, tempatnya yang bagus ombaknya besar,	2 (Netral)

tapi cukup dilihat dari jauh saja, tidak usah terlalu mendekati ombak, karena sering terjadi kecelakaan, pengunjung yang terseret ombak loh, karena mungkin kurang berhati hati. Untuk menuju pantai ini kami mengambil jalur arah balaikambang. Aksesnya lebih cepat, jalannya juga masih baru. Selain itu kita juga bisa menikmati pantai lepas sepanjang jalan.	
Kalo kita ke sana pas musim panas aja ..karena klo musim ujan air lautnya keruh ...jadi waktu snorkling ikannya ga kelihatan ...o iya jangan lupa bawa roti untuk makan ikan supaya naik ke atas....selamat berlibur	2 (Netral)

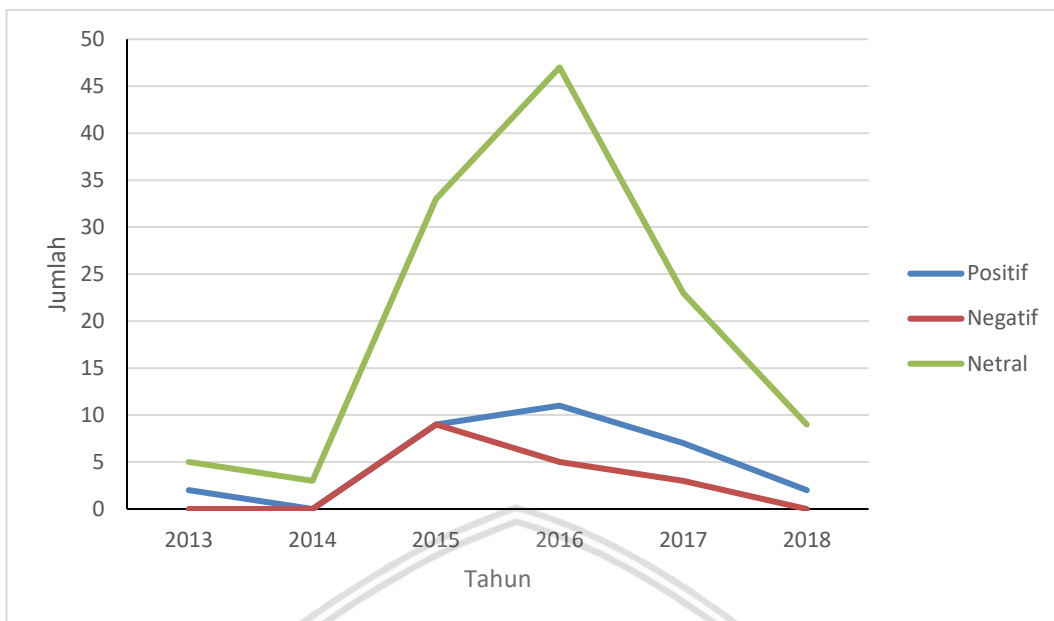
Pada Gambar 6.4 menunjukkan aspek Kebersihan menghasilkan 17 ulasan dengan sentimen negatif, 31 ulasan dengan sentimen positif, dan 120 ulasan dengan sentimen netral.



**Gambar 6.4 Frekuensi hasil klasifikasi terhadap aspek Kebersihan**

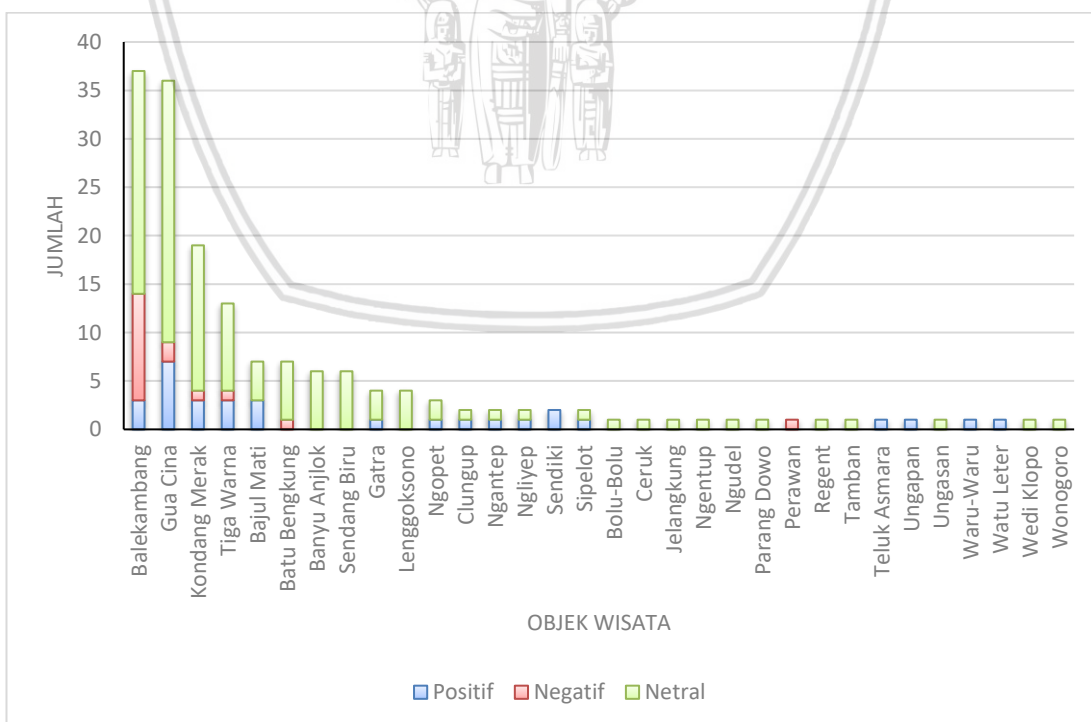
Pada Gambar 6.5 memperlihatkan tren dari ulasan terhadap aspek Kebersihan. Dari Gambar tersebut menunjukkan pada tahun 2014 hingga 2016 terjadi peningkatan ulasan positif terhadap aspek Kebersihan dengan puncaknya hingga 11 ulasan pada tahun 2016. Sedangkan untuk tahun 2016-2018 ulasan positif cenderung turun hingga hanya 2 ulasan positif untuk tahun 2018. Ulasan negatif memuncak pada tahun 2015 dengan total 9 ulasan negatif namun mengalami penurunan hingga pada 2017 hanya ada 3 ulasan negatif, dan 2018 dengan belum ada ulasan negatif.





**Gambar 6.5 Tren hasil klasifikasi terhadap aspek Kebersihan**

Gambar 6.6 menunjukkan jumlah ulasan berdasarkan nama objek wisata pantai Malang Selatan. Dari hasil klasifikasi menunjukkan pantai Balekambang, Gua Cina, dan Kondang Merak berturut-turut menjadi 3 teratas pantai yang memiliki ulasan dalam aspek Kebersihan dengan masing masing memiliki 37, 36, dan 19 ulasan.



**Gambar 6.6 Ulasan berdasarkan objek wisata terhadap aspek Kebersihan**

### 6.1.3 Aspek Keramaian

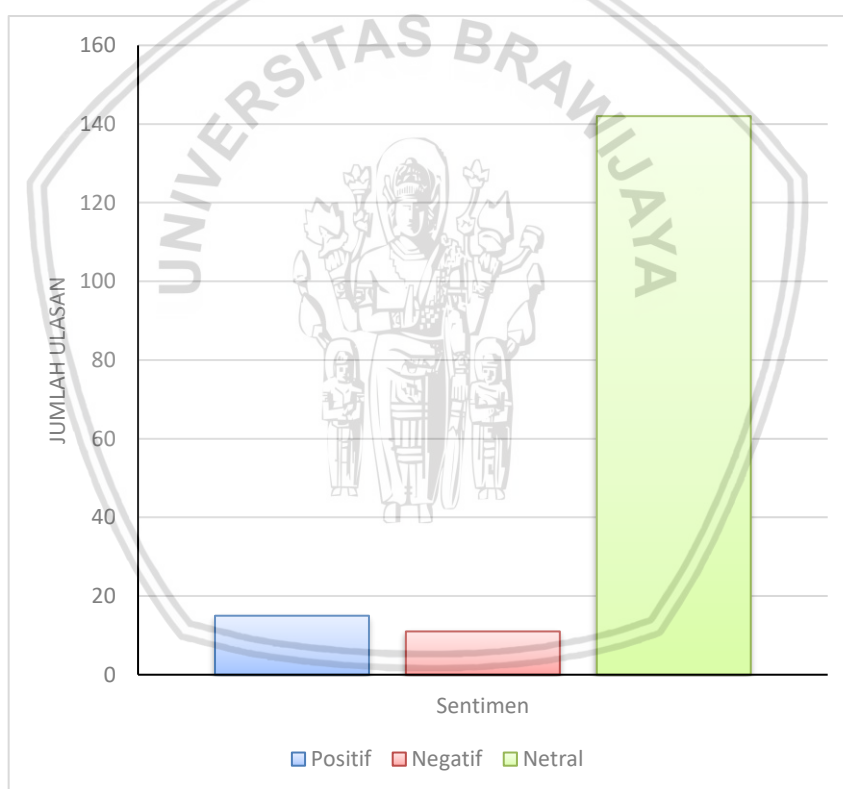
Aspek Keramaian meliputi kondisi tingkat keramaian pengunjung dari objek wisata pantai Malang Selatan. Dalam aspek Keramaian, *class* positif berisi ulasan yang menyatakan bahwa tingkat keramaian pengunjung rendah atau relatif sepi. Sedangkan *class* negatif berisi ulasan yang menyatakan tingkat keramaian pengunjung tinggi atau relatif ramai. Tabel 6.3 menunjukkan contoh teks ulasan yang berhasil diklasifikasikan ke dalam 3 sentimen yang ada yakni positif negatif dan netral.

**Tabel 6.3 Hasil klasifikasi aspek Keramaian**

Teks	Kelas
Salah satu wisata alternatif di Malang yang mulai ramai dikunjungi. Untuk berkunjung ke sini pengunjung dapat menggunakan perahu dari pantai Lenggoksono sekitar 20-30 menit. Dari jauh pengunjung sudah dapat melihat air terjun banyu anjlok yang airannya mengarah ke laut. Selain menikmati air terjun, pengunjung juga bisa menaiki tangga menuju ke atas untuk berenang di kolam alami yang terbentuk dari tumpahan air terjun	0 (Negatif)
Pantai yang cukup jauh dari kota Malang dan membutuhkan kendaraan pribadi untuk menjangkaunya. Pantai ini bagi saya tidak terlalu menarik bagi saya karena cukup ramai sehingga kurang bisa menikmatinya. Akankah lebih baik jika anda ke sini saat hari biasa untuk menikmatinya lebih puas.	0 (Negatif)
Jalan Menuju Pantai Jonggring Saloka cukup parah, rusak sekitar 6 km namun dapat ditempuh dengan mobil dan motor, kami tiba dipantai ceruk dan harus berjalan kaki melewati bukit sekitar 150 meter baru tiba dipantai Jonggring saloka, garis pantai cukup panjang dengan ada pulau kecil didepan pantai, dan banyak batu batu di pinggir pantai, sejak kami tiba hingga kami pulang sekitar 1,5 jam tidak ada 1 orang pun yang datang seakan pantai pribadi kami, pantai masih perawan dan ombak cukup besar, senang kami bermain disana	1 (Positif)
jalan menuju pantai ini adalah perempatan terakhir sebelum menuju ke balekambang anda ambil arah ke gua cina, kurang lebih 3 km akan ada petunjuk sebelah kanan pantai ngantep, pantai tidak terlalu banyak pengunjung namun cukup indah dengan air bening namun ombak agak besar, dipantai ini sudah dikelola ada tarif masuk 6500 per orang dan kendaraan roda 4 tarif 5000 rupiah	1 (Positif)
Pantai Ngantep terletak di Dusun Sukorejo, Desa Tumpakrejo, Kecamatan Gedangan, Kabupaten Malang, Jawa Timur. Pantai ini di kelola oleh pihak perhutani dan masyarakat sekitar Pantai Ngantep, jalan menuju Pantai Ngantep cukup mudah di lalui dengan kendaraan pribadi. Menikmati hamparan pasir putih yang halus ditambah dengan deburan ombak biru serta tempat yang	2 (Netral)

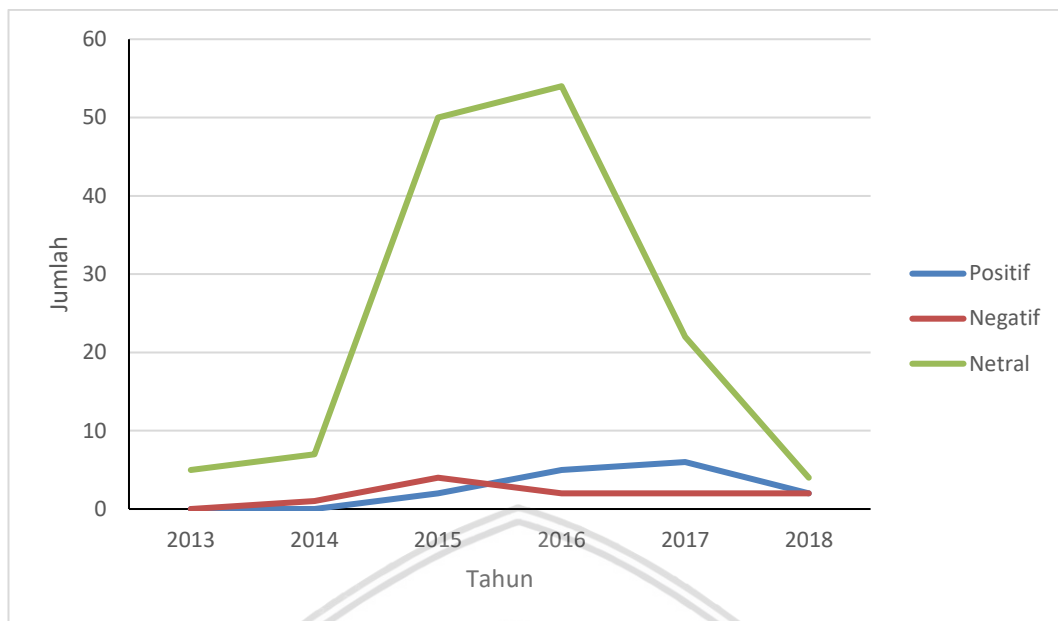
tenang akan menambah keinginan hati siapa saja untuk kembali lagi menyaksikan keindahan Pantai Ngantep.	
Perjalanan yg luar biasa dari malang ke pantai ini.. Betul betul sangat menguras tenaga ( tenaga mobil maksudnya ..hehe)dengan jalan yg penuh tanjakan ekstrim dan berliku seperti kehidupan...ada tiga lokasi favorit dipantai sini..pantai bolubolu,spot snorkeling dan banyu anjlok..harus menyeberang untuk mencapai lokasi tersebut..siapkan nyali dan doa untuk naek perahu disini..tapi dengan skill para pilot dan kopilot perahu yg sudah tinggi jam terbang semua itu bisa dilalui...sampai jumpa di pantai lenggoksono	2 (Netral)

Pada Gambar 6.7 menunjukkan aspek Keramaian menghasilkan 11 ulasan dengan sentimen negatif, 15 ulasan dengan sentimen positif, dan 142 ulasan dengan sentimen netral.



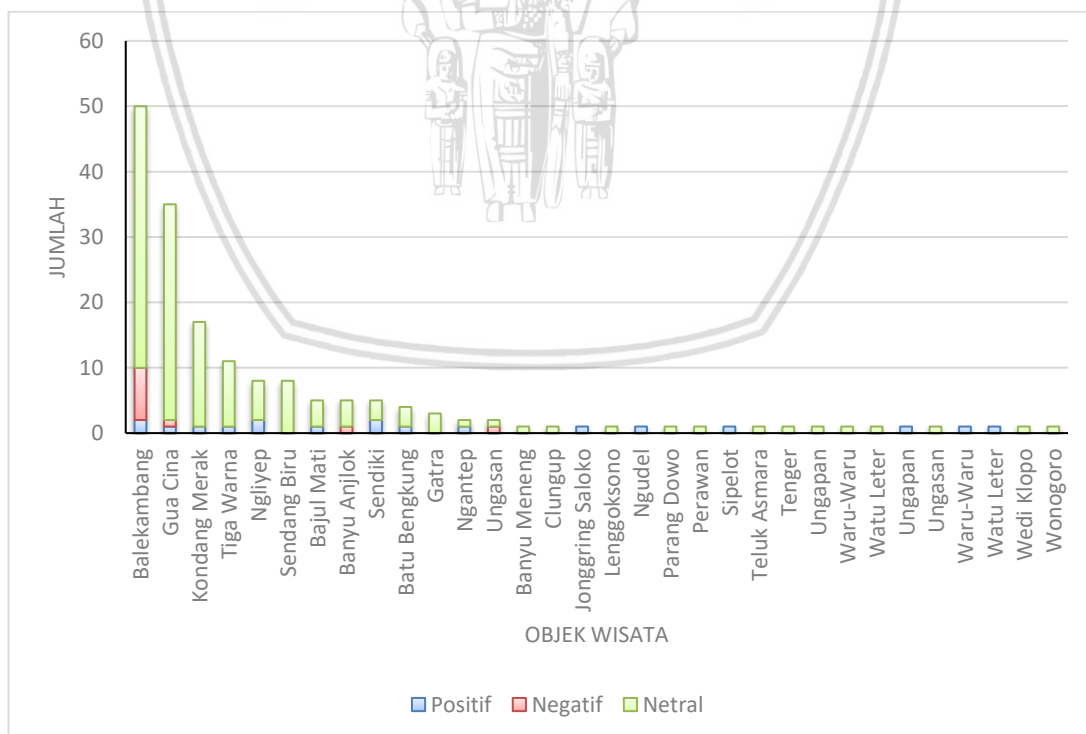
**Gambar 6.7 Frekuensi hasil klasifikasi terhadap aspek Keramaian**

Pada Gambar 6.8 memperlihatkan tren dari ulasan terhadap aspek Keramaian. Dari Gambar tersebut menunjukkan pada tahun 2013 hingga 2017 terjadi peningkatan ulasan positif terhadap aspek Keramaian dengan puncaknya hingga 6 ulasan pada tahun 2017. Sedangkan untuk tahun 2018 ulasan positif baru 2 ulasan positif. Ulasan negatif memuncak pada tahun 2015 dengan total 4 ulasan negatif namun mengalami penurunan hingga pada 2016 hanya ada 2 ulasan negatif, dan masih stagnan pada 2 ulasan negatif pada 2017 dan 2018.



**Gambar 6.8 Tren hasil klasifikasi terhadap aspek Keramaian**

Gambar 6.9 menunjukkan jumlah ulasan berdasarkan objek wisata pantai Malang Selatan. Dari hasil klasifikasi menunjukkan pantai Balekambang, Gua Cina, dan Kondang Merak berturut-turut menjadi 3 teratas pantai yang memiliki ulasan dalam aspek Keramaian dengan masing masing memiliki 50, 35, dan 17 ulasan.



**Gambar 6.9 Ulasan berdasarkan objek wisata terhadap aspek Keramaian**

#### 6.1.4 Aspek Akses Jalan

Aspek Akses Jalan meliputi kondisi tingkat kemudahan akses menuju objek wisata pantai Malang Selatan. Dalam aspek Akses Jalan, *class* positif berisi ulasan yang menyatakan bahwa akses jalan cenderung baik atau mudah. Sedangkan *class* negatif berisi ulasan yang menyatakan akses jalan cenderung buruk atau sulit. Tabel 6.4 menunjukkan contoh teks ulasan yang berhasil diklasifikasikan ke dalam 3 sentimen yang ada yakni positif negatif dan netral.

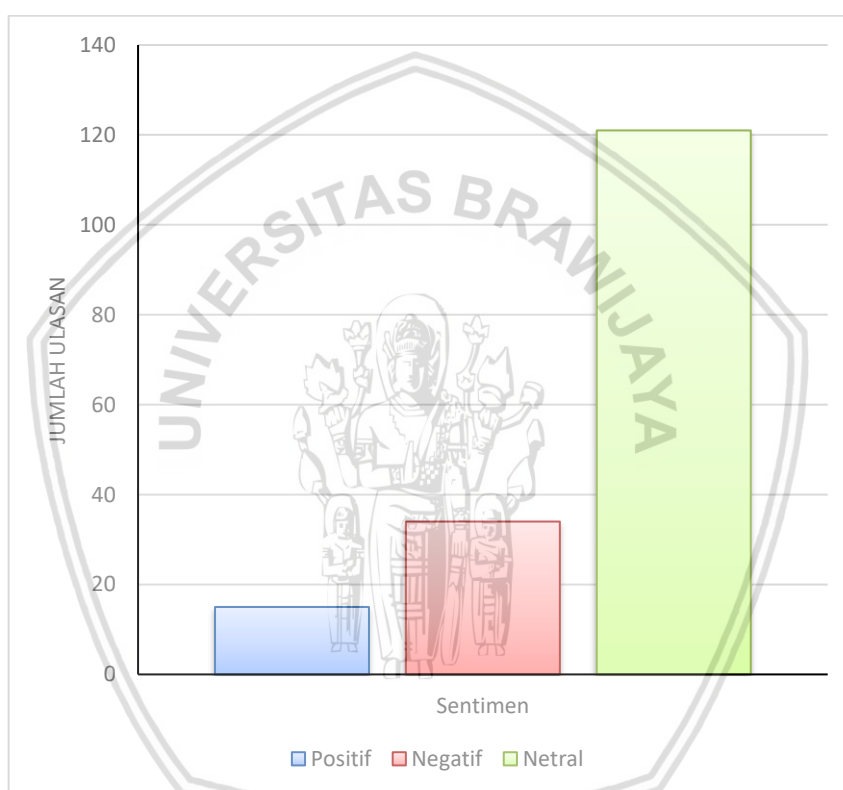
**Tabel 6.4 Hasil klasifikasi aspek Akses Jalan**

Teks	Kelas
Destinasi wisata pantai yang berbeda di Malang selatan dengan panorama yg indah laut yg tenang diantara gugusan pulau pulau kecil sangat cocok dan aman untuk berenang recomanded banget untuk dikunjungi. Cuma perjalanan dr malang kota ke lokasi jauh banget, masih harus jalan kaki 1.5 km lagi. Tapi keren banget pemandangannya.	0 (Negatif)
Pantai Balekambang ini terletak tidak terlalu jauh dari pusat kota Malang, hanya sekitar 1 jam perjalanan dengan mobil. Pasirnya memang tidak halus seperti bedak bayi, tapi airnya masih bening dan pemandangan pagi dan sore hari sangat bagus.	0 (Negatif)
Berbeda dengan pantai-pantai yang ada di daerah Malang Selatan, Pantai Ugasan memiliki kelebihan dari sisi pemandangan dan lokasinya. Di pantai tersebut, wisatawan dapat bermain air di muara yang menuju ke pantai. Selain airnya cukup dangkal, muara tersebut juga tidak berbahaya (tidak ada binatang buas). Bagi yang membawa anak kecil, dapat terpuaskan dengan bermain air di muara terebut. Tapi jangan sekali-kali bermain di Pantainya, karena Ombaknya sangat besar dan berbahaya. Cukup dinikmati deburan ombaknya saja. Bagi yang ingin berkunjung ke Pantai Ugasan, lokasinya bedekatan dengan Pantai Sendang Biru, Pantai Goa Cina, Pantai Bajul Mati dll (satu jalan bisa menuju ke banyak pantai, sudah Hot Mix pula jalannya hehehehehe). Dengan harga tiket yang murah, kita dapat terpuaskan dengan keindahan Pantai Ugasan ini. Selamat berkunjung ke Pantai Ugasan.....	1 (Positif)
lokasi akses masuk sekarang cukup mudah, sudah ada jalan rata. tempat ini sangat cocok untuk wisata keluarga, pasir dan pantainya bersih, namun jangan sekali kali mandi, masih banyak karang tajam. tempat ini juga cocok untuk berkemah dan touring motor. ada karang bolong yang konon pernah menjadi tempat pertapaan	1 (Positif)
Menyenangkan saat melihat pemandangan disana, hanya kurang fasilitas seperti tempat duduk, spot-spot foto yang menarik seperti tulisan atau berupa hiasan disekitaran pantai. ditengah pantai ada pulau kecil pulau Bantengan, Pulau Goa Cina dan Pulau Nyonya.	2 (Netral)



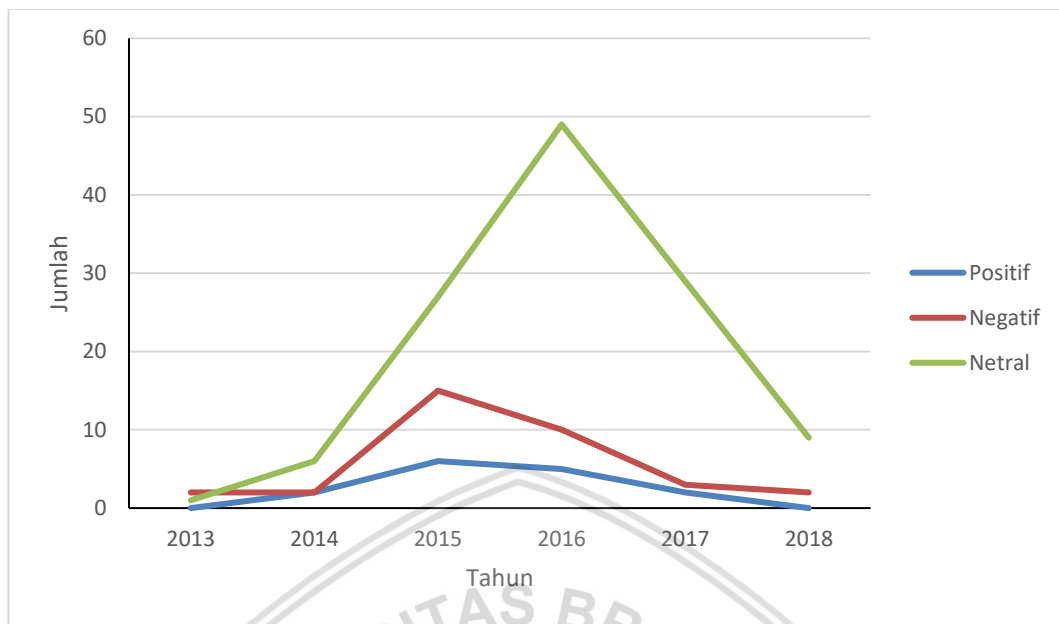
Udarasegar berpadu dg debur ombak berair biru, bikin fikiran fresh. Kuliner (seafood) jg lumayan lezat dan lengkap	
Kalau kita mau ke sendang biru, beberapa kilometer sebelumnya ada pertigaan ke arah kiri, disana masuk sekitar 5 km ada 2 pantai yaitu pantai sendiki dan tamban, pantai tamban adalah pantai nelayan, banyak perahu nelayan disana, pemandangan tidak ada yang spesial disana	2 (Netral)

Pada Gambar 6.10 menunjukkan aspek Akses Jalan menghasilkan 34 ulasan dengan sentimen negatif, 15 ulasan dengan sentimen positif, dan 121 ulasan dengan sentimen netral.



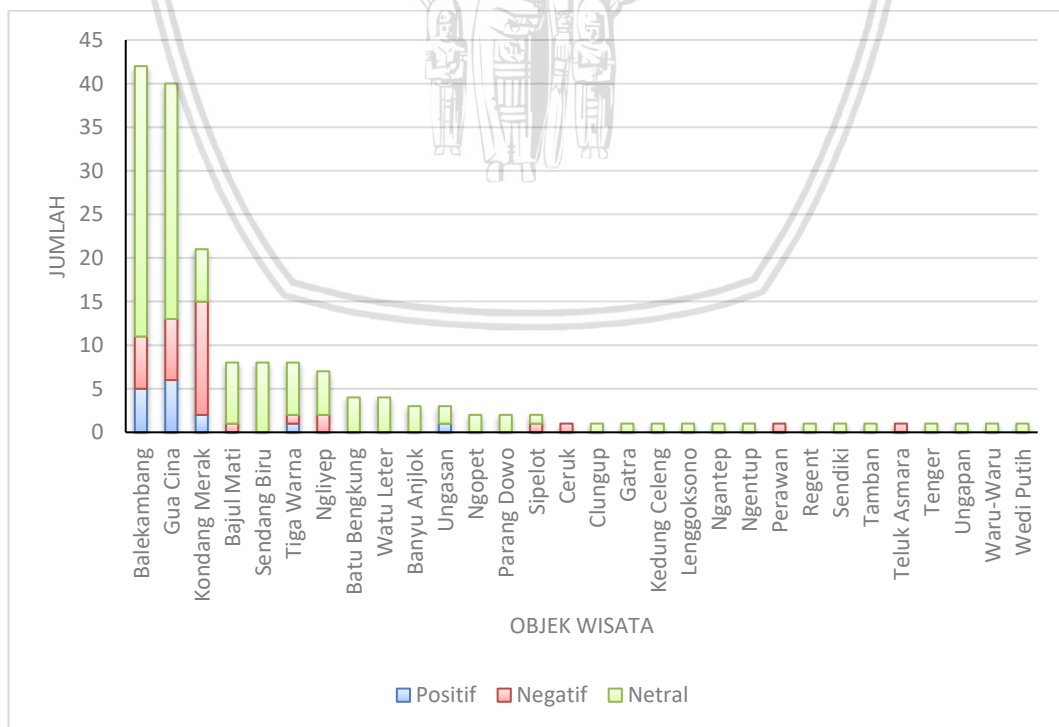
**Gambar 6.10 Frekuensi hasil klasifikasi terhadap aspek Akses Jalan**

Pada Gambar 6.11 memperlihatkan tren dari ulasan terhadap aspek Akses Jalan. Dari Gambar tersebut menunjukkan pada tahun 2013 hingga 2015 terjadi peningkatan ulasan positif terhadap aspek Akses Jalan dengan puncaknya hingga 6 ulasan pada tahun 2015. Sedangkan untuk tahun 2016-2018 ulasan positif baru mengalami penurunan dari 5 ulasan positif pada tahun 2016, 2 ulasan positif pada tahun 2017, dan belum ada ulasan positif pada tahun 2018. Ulasan negatif memuncak pada tahun 2015 dengan total 15 ulasan negatif namun mengalami penurunan hingga pada 2016 ada 10 ulasan negatif, dan pada 2017 menjadi 3 ulasan negatif. Sementara pada tahun 2018 sudah ada 2 ulasan negatif untuk aspek Akses Jalan.



**Gambar 6.11 Tren hasil klasifikasi terhadap aspek Akses Jalan**

Gambar 6.12 menunjukkan jumlah ulasan berdasarkan objek wisata pantai Malang Selatan. Dari hasil klasifikasi menunjukkan pantai Balekambang, Gua Cina, dan Kondang Merak berturut-turut menjadi 3 teratas pantai yang memiliki ulasan dalam aspek Akses Jalan dengan masing masing memiliki 42, 40, dan 21 ulasan.



**Gambar 6.12 Ulasan berdasarkan objek wisata terhadap aspek Akses Jalan**

### 6.1.5 Aspek Ombak

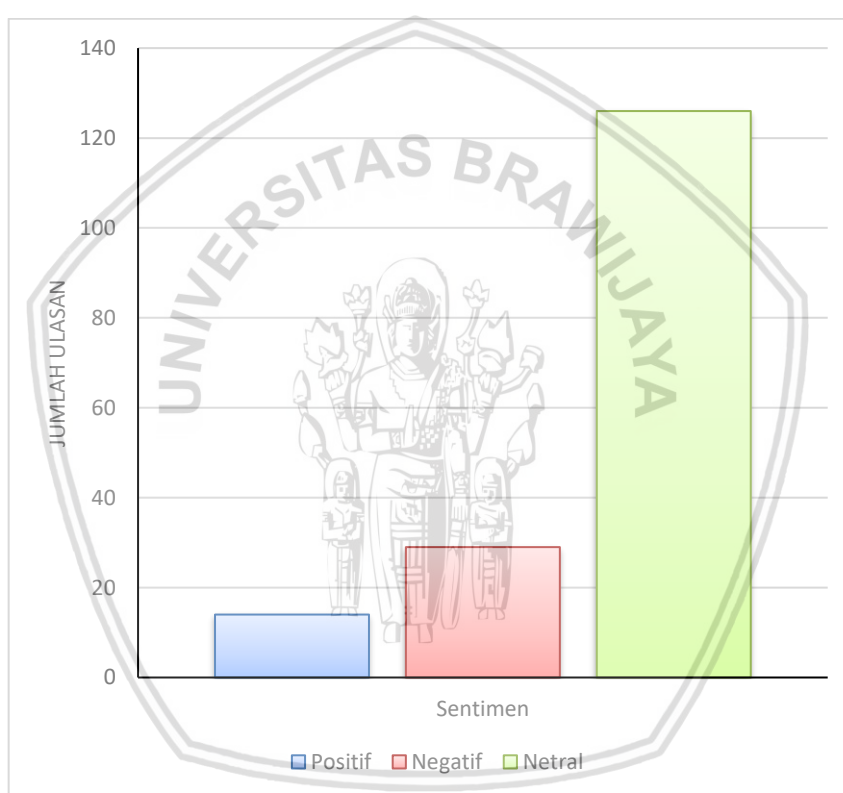
Aspek Ombak meliputi kondisi ombak dan perairan yang ada pada objek wisata pantai Malang Selatan. Dalam aspek Ombak, *class* positif berisi ulasan yang menyatakan bahwa ombak rendah atau perairan cenderung tenang dan tidak berbahaya. Sedangkan *class* negatif berisi ulasan yang menyatakan bahwa ombak tinggi atau perairan cenderung berbahaya. Tabel 6.4 menunjukkan contoh teks ulasan yang berhasil diklasifikasikan ke dalam 3 sentimen yang ada yakni positif negatif dan netral.

**Tabel 6.5 Hasil klasifikasi aspek Ombak**

Teks	Kelas
Bajul Mati - cukup banyak tidak diketahui bahkan untuk penduduk setempat adalah pantai yang terletak di deretan pantai Goa Cina dan Balekambang antara pantai ini menawarkan pantai alami alam dan udara yang bersih, beberapa - jika - wisatawan lain, maka ini akan menjadi tempat yang sempurna untuk seseorang yang sembilan tempat yang unik tidak akan mengganggu liburan Anda lebih baik anda tidak berenang di sini karena ombaknya yang tinggi	0 (Negatif)
Pantai ini kalo tidak salah berjarak 2 jam setengah dari pusat kota Malang. Tidak seperti pantai-pantai yang lain yang berhawa panas, pantai ini jauh lebih adem karena letaknya di sekitar pegunungan. Pantainya sangat cantik dengan air yang jernih dan pasir putih tapi sayangnya tidak dibolehkan untuk berenang karena ombaknya yang lumayan besar.	0 (Negatif)
Jujur ini adalah pantai favorit saya di malang. Cenderung sepi karena hanya yg mencarter perahu yg bisa ke sini. Bersih. Ombaknya tidak tinggi jadi anda bisa berenang selama masih di tepi pantai. Ada karang2 bagus untuk berfoto juga. Kalau lapar ada warung yg jual makanan dengan harga yg masih terjangkau. Jangan lupa banyak minum air putih agar tak dehidrasi	1 (Positif)
Merupakan pantai yang relatif paling aman untuk bermain riak kecil disepanjang pantai Kondangmerak. Ombak yang pecah ditengah karena gugusan karang, arus yang searah, menjadikan Pantai Kondangmerak sebagai pantai yg relatif paling aman dibanding pantai wisata lainnya. Mendidik, karena banyak spot transplantasi terumbu karang... Alami..., semoga dengan akses jalan masuk yg berat, menjadikan area pantai Kondangmerak senantiasa terjaga kealamiannya.	1 (Positif)
Pantai Lenggoksongo, berada di wilayah Dampit Kabupaten Malang, pantainya cukup indah, dengan bibir pantai yang berbatu-batu, menambah eksotisnya pantai. Disekitar pantai terdapat banyak penjual makanan, pantai ini merupakan kampung nelayan, mereka sangat ramah. Pantai Lenggoksongo merupakan pintu masuk menuju ke Air Terjun Banyu Anjlok, yang terletak sekitar 3	2 (Netral)

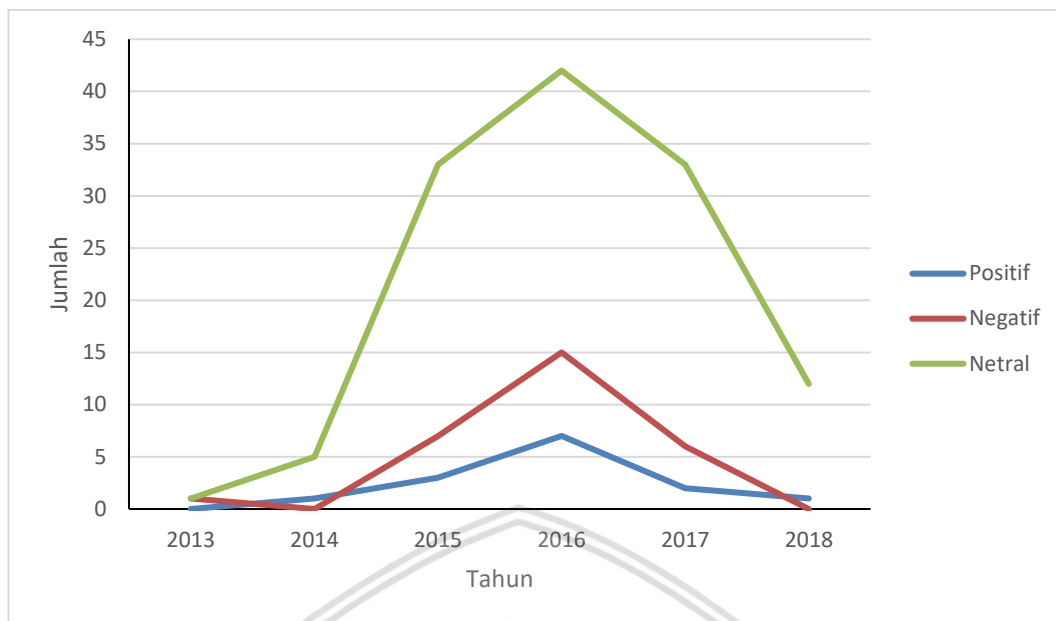
Km. Tempat wisata ini juga dilengkapi Toilet umum dan tempat Parkir. Sangat cocok untuk wisata keluarga.	
tiba disana pagi hari memang sangat menyenangkan, udara masih segar menusuk hidung dan suara ombak yang keren, sayangnya kebersihan kurang terjaga meski sudah ada tempat sampah dan papan nama untuk menjaga kebersihan. Tapi sangat menyenangkan bermain ditepi pantai bersama sama teman gereja.semoga bisa kembali lagi kesana	2 (Netral)

Pada Gambar 6.13 menunjukkan aspek Ombak menghasilkan 29 ulasan dengan sentimen negatif, 14 ulasan dengan sentimen positif, dan 126 ulasan dengan sentimen netral.



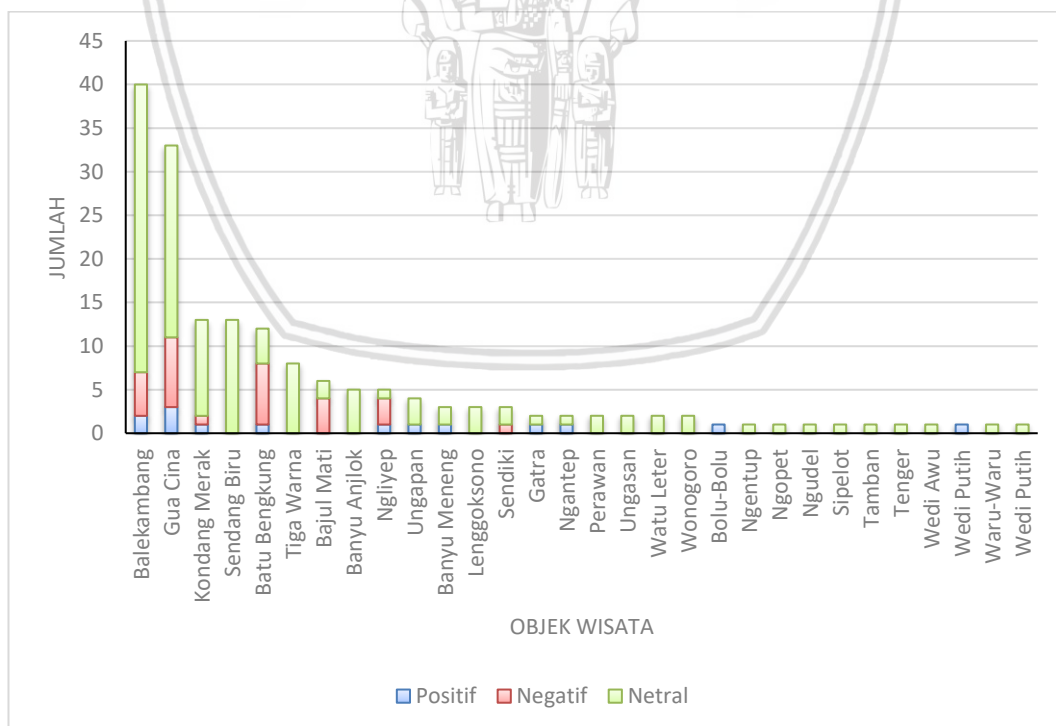
**Gambar 6.13 Frekuensi hasil klasifikasi terhadap aspek Ombak**

Pada Gambar 6.14 memperlihatkan tren dari ulasan terhadap aspek Ombak. Dari Gambar tersebut menunjukkan pada tahun 2013 hingga 2016 terjadi peningkatan ulasan positif terhadap aspek Ombak dengan puncaknya hingga 7 ulasan pada tahun 2016. Sedangkan untuk tahun 2017-2018 ulasan positif baru 2 ulasan positif pada tahun 2017, dan 1 ulasan positif pada tahun 2018. Ulasan negatif memuncak pada tahun 2016 dengan total 15 ulasan negatif namun mengalami penurunan hingga pada 2017 ada 6 ulasan negatif, dan Sementara pada tahun 2018 belum ada ulasan negatif untuk aspek Ombak.



**Gambar 6.14 Tren hasil klasifikasi terhadap aspek Ombak**

Gambar 6.15 menunjukkan jumlah ulasan berdasarkan objek wisata pantai Malang Selatan. Dari hasil klasifikasi menunjukkan pantai Balekambang, Gua Cina, dan Kondang Merak berturut-turut menjadi 3 teratas pantai yang memiliki ulasan dalam aspek Ombak dengan masing masing memiliki 40, 33, dan 13 ulasan.



**Gambar 6.15 Ulasan berdasarkan objek wisata terhadap aspek Ombak**



## 6.2 Pengujian hasil klasifikasi sentimen

Pengujian hasil klasifikasi sentimen dilakukan dengan melakukan 4 parameter pengujian yakni *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. *Source code* 6.1 menjelaskan proses pengujian dengan memanfaatkan *module Scikit-learn* pada Bahasa Pemrograman Python.

```

1  from sklearn.metrics import classification_report
2  from sklearn.metrics import confusion_matrix
2  X_train='datalatih'
3  X_test='datauji'
4  y_train='labellatih'
5  y_test='labeluji'
6  svm_model_linear = LinearSVC(multi_class='ovr').fit(X_train,
7  y_train)
7  pred = svm_model_linear.predict(X_test)
8  accuracy = svm_model_linear.score(X_test, y_test)
9  print(confusion_matrix(y_test, pred))
10 print(classification_report(y_test, pred))
11 print(accuracy)
12 print(pred)

```

**Source code 6.1 Source code pengujian klasifikasi sentimen**

Tabel 6.6 menunjukkan hasil pengujian terhadap hasil klasifikasi sentimen. Nilai *Accuracy* tertinggi didapat pada model klasifikasi untuk aspek Keramaian dengan nilai 0,92. Nilai akurasi terendah didapat pada model klasifikasi untuk aspek Umum dan Akses Jalan dengan nilai 0,85. Nilai *Precision* tertinggi pada model klasifikasi untuk aspek Keramaian dengan nilai 0,93 dan terendah pada model klasifikasi untuk aspek Umum dengan nilai 0,75. Nilai *Recall* tertinggi dihasilkan oleh model klasifikasi untuk aspek Keramaian dengan nilai 0,92 dan terendah dihasilkan oleh model klasifikasi untuk aspek Umum dan Akses Jalan dengan nilai 0,85. Nilai *F1-Score* tertinggi pada model klasifikasi untuk aspek Keramaian dengan nilai 0,91 dan terendah pada model klasifikasi untuk akses Umum yakni 0,79. Bila dilakukan rata-rata terhadap nilai pengujian hasil klasifikasi sentimen maka hasilnya adalah sebagai berikut: *Accuracy* sebesar 0,87, *Precision* sebesar 0,85, *Recall* sebesar 0,87, dan *F1-Score* sebesar 0,85.

**Tabel 6.6 Hasil pengujian akurasi**

Aspek	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Umum	0,85	0,75	0,85	0,79
Kebersihan	0,87	0,87	0,87	0,86
Keramaian	0,92	0,93	0,92	0,91
Akses Jalan	0,85	0,84	0,85	0,83
Ombak	0,87	0,87	0,88	0,86
Rata-rata	0,87	0,85	0,87	0,85

## 6.3 Visualisasi *Dashboard*

Tahapan visualisasi memanfaatkan *tool Plotly*. *Plotly* adalah untuk membuat, mengubah, dan berbagi visualisasi data dalam platform *Web* (*Plotly*, 2017).

### 6.3.1 *Dashboard Analisis Sentimen*

Gambar 6.16 pada halaman 72 menunjukkan hasil visualisasi *Dashboard Analisis Sentimen*. Dalam *Dashboard* tersebut menampilkan hasil klasifikasi sentimen opini pelanggan terhadap pariwisata pantai Malang Selatan. Dalam visualisasi tersebut menampilkan persentase sentimen dari 3 kelas sentimen yakni positif, negatif, dan netral untuk 5 aspek yang ada. Data ditampilkan dalam bentuk *Bar chart*. Sentimen positif ditunjukkan oleh *bar* berwarna biru, sentimen negatif ditunjukkan oleh *bar* berwarna jingga, dan sentimen netral ditunjukkan dengan *bar* berwarna hijau. Pada bagian kanan terdapat *Legend* yang dapat digunakan untuk mengatur tampilan data berdasarkan kelas sentimennya. Pada bagian atas terdapat 4 *button* untuk dapat berpindah ke *Dashboard* lain.

### 6.3.2 *Dashboard trend sentimen*

Gambar 6.17 pada halaman 73 menunjukkan hasil visualisasi *Dashboard trend sentimen*. *Dashboard trend sentimen* memvisualisasikan perubahan sentimen opini pelanggan terhadap aspek pariwisata pantai Malang Selatan dari tahun 2013 hingga tahun 2018. Dalam *Dashboard* tersebut menampilkan data persentase dari 3 kelas sentimen untuk setiap aspek dari tahun 2013 hingga tahun 2018. Data dikelompokkan berdasarkan aspek. Untuk mengatur tampilan data bisa menggunakan kotak abu-abu yang didalamnya terdapat *button* untuk menampilkan data berdasarkan aspek. Kemudian adanya legenda untuk menerangkan penggunaan warna yang menjadi representasi masing-masing sentimen. Pada bagian atas terdapat 4 *button* untuk dapat berpindah ke *Dashboard* lain.

### 6.3.3 *Dashboard objek wisata*

Gambar 6.18 pada halaman 74 menunjukkan hasil visualisasi *Dashboard objek wisata*. *Dashboard objek wisata* merupakan *Dashboard* yang menampilkan pemeringkatan objek pariwisata pantai Malang Selatan terhadap 5 aspek yang ada. *Dashboard* menampilkan data dalam bentuk *radar chart* dengan 5 sumbu yang menunjukkan 5 aspek yang ada. Data yang dijadikan *input* adalah sentimen positif dari semua aspek yang dimiliki oleh suatu objek wisata pantai. Dalam visualisasi ini menampilkan 17 objek wisata pantai Malang Selatan. Ketujuh belas objek wisata tersebut dapat ditampilkan maupun dibandingkan dengan objek wisata lainnya dengan cara mengatur penyajian data melalui *button* untuk setiap objek wisata yang terletak di bagian kanan *radar chart*.

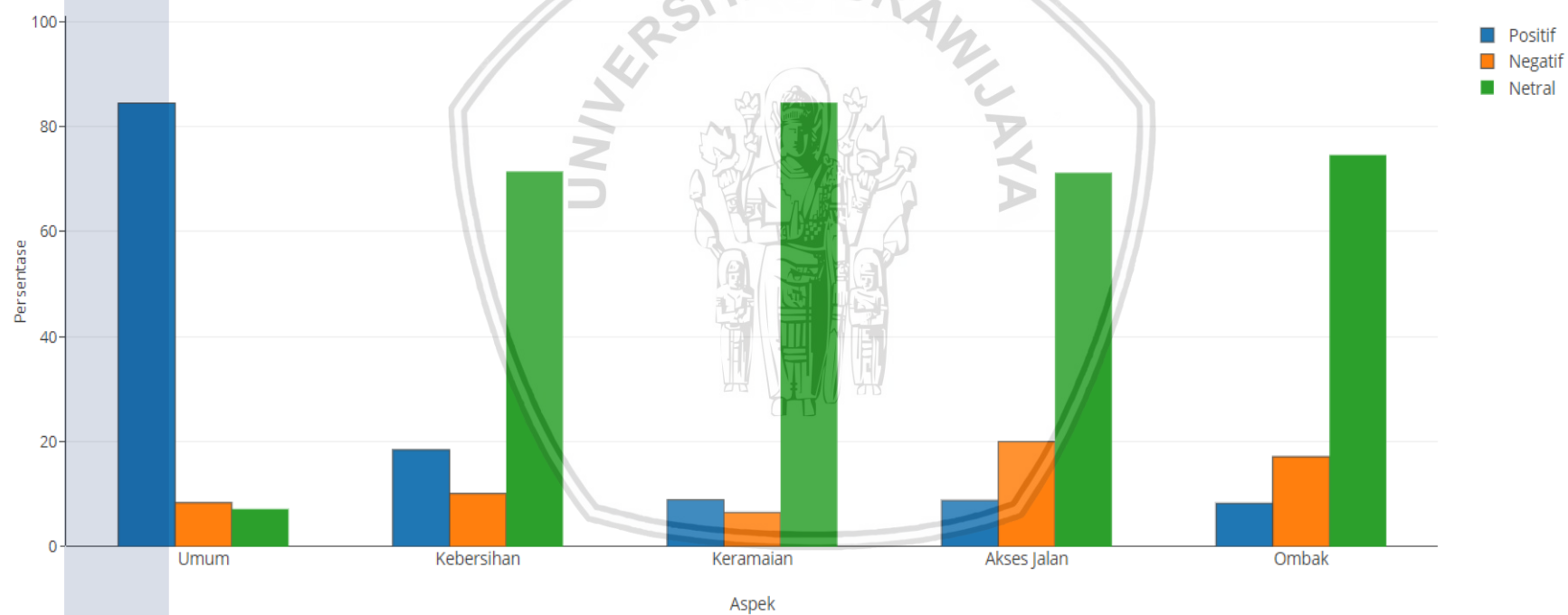
#### 6.3.4 *Dashboard* aspek

Untuk tampilan visualisasi *Dashboard* aspek terlampir pada Lampiran C. *Dashboard* aspek memvisualisasikan data sentimen dari tahun 2013 hingga 2018 dari setiap objek wisata pantai Malang Selatan untuk setiap aspek. *Dashboard* berisi 5 *bar chart* yang masing-masing mewakili 5 aspek pariwisata pantai Malang Selatan. Untuk penyajian data dapat diatur melalui *button* pada sebelah kanan dari *bar chart*. Data dikelompokkan berdasar objek pariwisata, sehingga akan menampilkan perubahan sentimen dari suatu objek pariwisata dari tahun ke tahun terhadap suatu aspek.



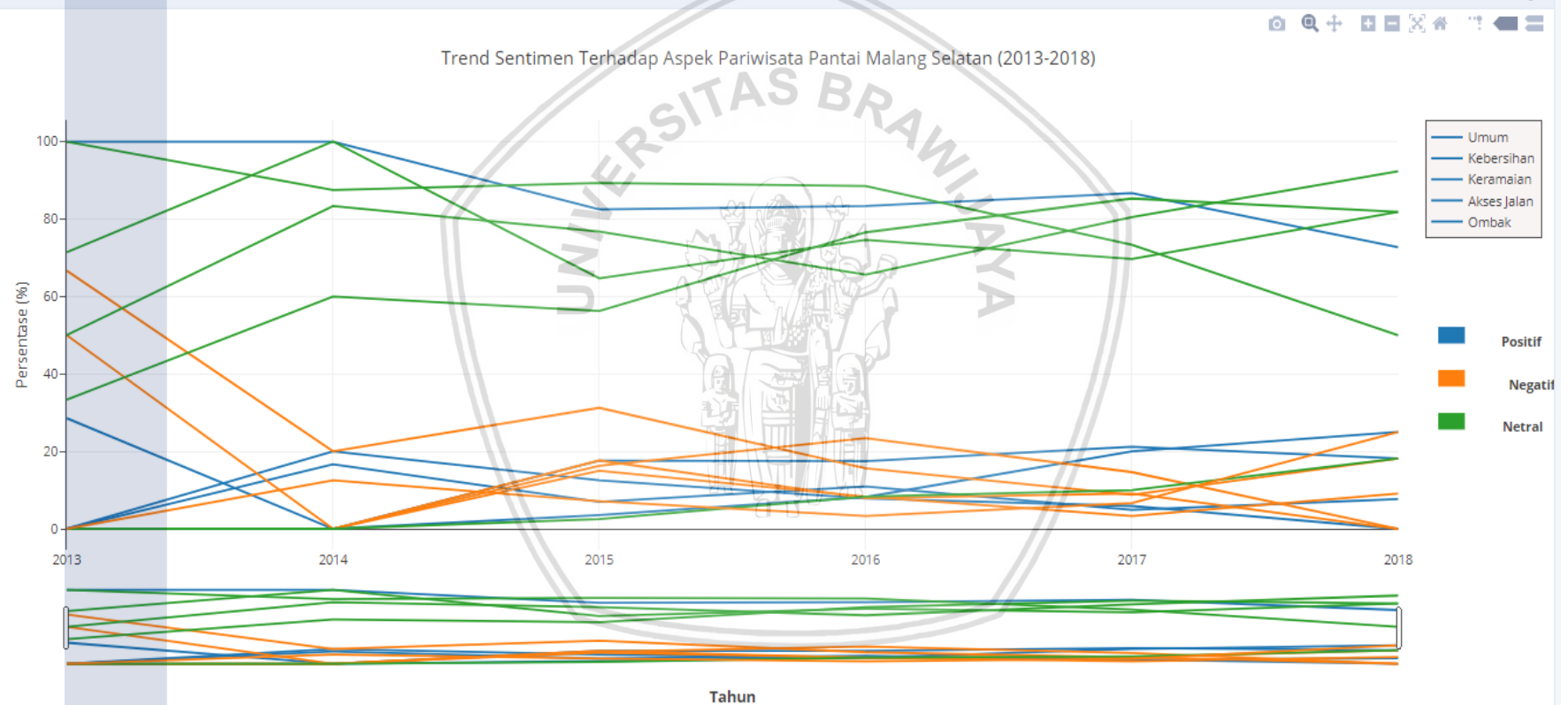
Dashboard Analisis Sentimen

Hasil Analisis Sentimen Opini Publik Terhadap Pariwisata Pantai Malang Selatan (2013-2018)

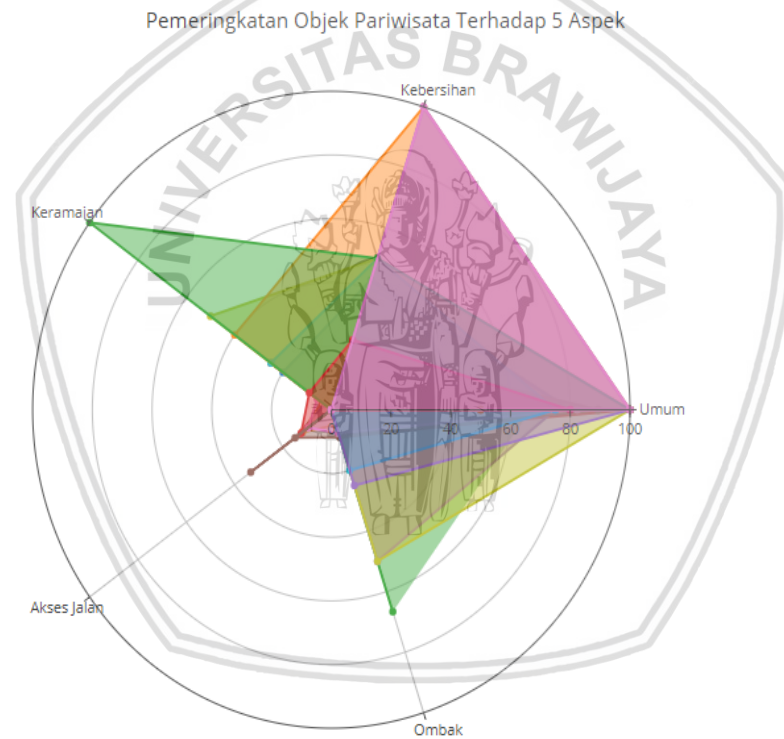


Gambar 6.16 Dashboard Analisis Sentimen

Dashboard Trend



Gambar 6.17 Dashboard Trend Sentimen



Gambar 6.18 Dashboard Objek Wisata



## 6.4 Analisis hasil *Usability Testing Dashboard*

*Usability Testing* dilakukan terhadap hasil visualisasi *Dashboard* dengan tujuan untuk menilai kualitas ketermudahan penggunaan *Dashboard* oleh pengguna. Bangor et al. (2008) mendefinisikan *usability* sebagai sejauh mana kualitas yang dimiliki sistem dalam segi kemudahan segi kemudahan untuk dipelajari, kemudahan untuk digunakan, dan mendorong pengguna untuk menggunakan sistem sebagai alat positif dalam menyelesaikan tugas. Metode yang digunakan dalam *Usability Testing* yakni *System Usability Scale* (SUS). Sauro (2011) Menyatakan bahwa kelebihan metode ini adalah dapat digunakan pada sampel yang jumlahnya sedikit. SUS menggunakan 10 pertanyaan dalam kuesioner. Untuk melakukan perhitungan nilai SUS dilakukan dengan cara memberi bobot antara 0 sampai dengan 4 dengan aturan sebagai berikut :

1. Item dengan nomor ganjil merupakan pernyataan positif yang nilainya didapat dengan cara posisi skala – 1.
2. Item dengan nomor genap merupakan pernyataan negatif yang nilainya didapat dengan cara 5 – posisi skala.
3. Hasil akumulasi pertanyaan positif dan negatif kemudian dikalikan 2,5.

Tabel 6.7 menjelaskan hasil *Usability Testing* dengan menggunakan SUS yang telah dilakukan oleh perwakilan pihak Dinas Pariwisata dan Kebudayaan Kabupaten Malang. Dinas Pariwisata dan Kebudayaan Kabupaten Malang diwakili oleh pejabat terkait dalam bidang pengembangan Sistem Informasi yakni Bapak Gede Arya. Kuesioner terlampir.

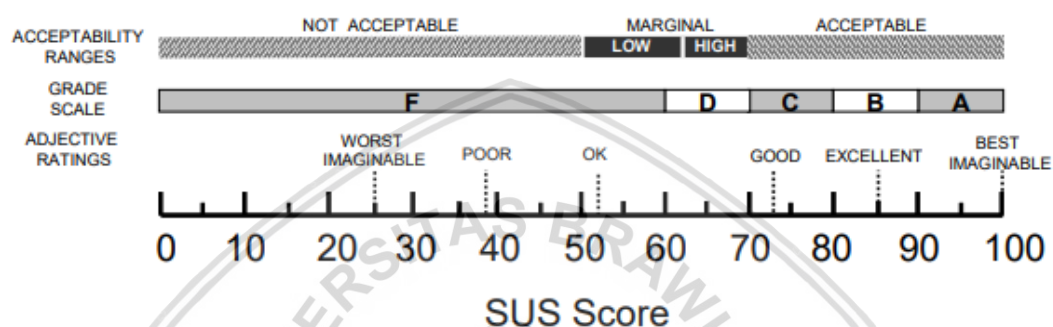
**Tabel 6.7 Hasil penilaian *System Usability Scale***

No.	Pertanyaan	Jawaban
1	Saya pikir saya akan sering menggunakan sistem ini	4
2	Saya rasa sistem ini terlalu kompleks	2
3	Saya kira sistem ini mudah untuk digunakan	4
4	Saya pikir saya akan membutuhkan bantuan teknis untuk dapat menggunakan sistem ini	4
5	Saya kira ada berbagai fungsi dalam sistem ini terintegrasi dengan baik	4
6	Saya pikir ada terlalu banyak ketidaksesuaian dalam sistem ini	3
7	Saya bisa bayangkan bahwa kebanyakan orang akan mudah mempelajari dan menggunakan sistem ini	4
8	Saya kira sistem ini sangat rumit untuk digunakan	2
9	Saya merasa percaya diri dalam menggunakan sistem ini	4
10	Saya perlu belajar banyak hal sebelum saya bisai memulai menggunakan sistem ini	4

Untuk mendapat nilai SUS dilakukan perhitungan sebagai berikut:

$$\begin{aligned} SUS \text{ Score} &= ((4-1)+(5-2)+(4-1)+(5-4)+(4-1)+(5-2)+(4-1)+(5-2)+(4-1)+(4-1)) * 2,5 \\ &= (3+3+3+1+3+3+3+3+3+3) * 2,5 \\ &= 70 \end{aligned}$$

Bangor et al. (2009) memberikan cara untuk dapat mengkategorikan hasil dari *Usability Testing* dengan menggunakan SUS. Gambar 6.19 menunjukkan bagaimana mengelompokkan hasil dari *SUS Score*. *SUS Score* bisa dikelompokkan dalam 3 parameter yakni *Acceptability Ranges*, *Grade Scale*, dan *Adjective Ratings*.



**Gambar 6.19 Skala pemeringkatan *SUS Score***

Sumber: Bangor et al. (2009)

Berdasarkan perhitungan yang telah dilakukan maka *SUS Score* untuk *Dashboard* sebesar 70 sehingga dapat disimpulkan bahwa *Dashboard* tergolong ke dalam kategori *Acceptable* dan dalam *rating "Good"* yang artinya *Dashboard* dapat diterima dengan baik oleh Dinas Pariwisata dan Kebudayaan Kabupaten Malang.

## BAB 7 PENUTUP

### 7.1 Kesimpulan

Setelah melakukan pembahasan penelitian yang telah ada, maka diperoleh kesimpulan-kesimpulan seperti berikut ini :

1. Teknik *Web Scraping* bisa digunakan untuk mendapatkan data opini pelanggan yang dibutuhkan untuk melakukan Analisis Sentimen. Hal tersebut dapat dibuktikan dengan mampunya teknik *Web Scraping* yang diimplementasikan untuk mendapatkan data ulasan dari situs TripAdvisor. Variabel data ulasan yang didapatkan dari situs TripAdvisor berupa judul ulasan, tanggal pemberian ulasan, pemberi ulasan, teks ulasan serta objek pariwisata yang diulas. Data yang didapatkan dari penerapan *Web Scraping* berupa data opini dalam bentuk ulasan terhadap pariwisata pantai Malang Selatan dengan total 674 data ulasan berbahasa Indonesia untuk 43 objek wisata pantai Malang Selatan dalam rentang waktu tahun 2013 hingga tahun 2018.
2. Metode *Support Vector Machine* (SVM) yang dikombinasikan dengan *Term Frequency-Inverse Document Frequency* (TF-IDF) dapat menjadi satu pilihan dalam menyelesaikan permasalahan klasifikasi sentimen dalam kerangka kerja Analisis Sentimen. Hal tersebut dibuktikan dengan nilai hasil *Term Weighting* TF-IDF untuk setiap data opini, beserta nilai label untuk masing-masing aspek dapat diklasifikasikan dengan menggunakan metode SVM untuk melakukan klasifikasi sentimen untuk setiap aspek. Klasifikasi sentimen menggunakan SVM mampu mengklasifikasikan sentimen untuk aspek Umum dengan menghasilkan 14 ulasan dengan sentimen negatif, 141 ulasan dengan sentimen positif, dan 12 ulasan dengan sentimen netral dengan nilai akurasi 85%. Untuk aspek Kebersihan berhasil mengklasifikasikan 17 ulasan dengan sentimen negatif, 31 ulasan dengan sentimen positif, dan 120 ulasan dengan sentimen netral dengan nilai akurasi 87%. Klasifikasi untuk aspek Keramaian menghasilkan 11 ulasan dengan sentimen negatif, 15 ulasan dengan sentimen positif, dan 142 ulasan dengan sentimen netral dengan nilai akurasi 92%. Klasifikasi untuk aspek Akses Jalan 34 ulasan dengan sentimen negatif, 15 ulasan dengan sentimen positif, dan 121 ulasan dengan sentimen netral dengan nilai akurasi 85%. Klasifikasi aspek Ombak menghasilkan 29 ulasan dengan sentimen negatif, 14 ulasan dengan sentimen positif, dan 126 ulasan dengan sentimen netral dengan nilai akurasi 87%.
3. Dari hasil implementasi metode *Support Vector Machine* kemudian dilakukan pengujian terhadap hasil klasifikasi sentimen untuk masing-masing aspek. Parameter pengujian berupa *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Pengujian terhadap hasil klasifikasi sentimen menghasilkan rata-

rata yang baik untuk *Accuracy* sebesar 87%, *Precision* sebesar 85%, *Recall* sebesar 87%, dan *F1-Score* sebesar 85%.

4. Proses *Usability Testing* terhadap *Dashboard* dengan menggunakan metode *System Usability Testing* (SUS) menghasilkan nilai yang tinggi yakni 70 yang memasukkan *Dashboard* sebagai sistem dengan kategori *Acceptable*. Visualisasi hasil Analisis Sentimen ditampilkan secara interaktif untuk pengambilan keputusan strategis Dinas Pariwisata dan Kebudayaan Kabupaten Malang. Proses visualisasi menghasilkan 4 visualisasi *Dashboard* yakni *Dashboard* Analisis Sentimen, *Dashboard trend* sentimen, *Dashboard* objek wisata, dan *Dashboard* aspek. *Dashboard* Analisis Sentimen menampilkan hasil klasifikasi sentimen opini pelanggan terhadap pariwisata pantai Malang Selatan. *Dashboard trend* sentimen memvisualisasikan perubahan sentimen opini pelanggan terhadap aspek pariwisata pantai Malang Selatan dari tahun 2013 hingga tahun 2018. *Dashboard* objek wisata adalah *Dashboard* yang menampilkan pemeringkatan objek pariwisata pantai Malang Selatan terhadap 5 aspek yang ada. *Dashboard* aspek memvisualisasikan data sentimen dari tahun 2013 hingga 2018 dari setiap objek wisata pantai Malang Selatan untuk setiap aspek.

## 7.2 Saran

Berikut adalah saran perbaikan yang dapat dilakukan pada penelitian selanjutnya :

1. Menambahkan sumber data opini pelanggan dari sumber data lain seperti Media Sosial, Google Reviews, dan platform lain sejenisnya yang memuat data opini pelanggan pariwisata Pantai Malang Selatan. Penambahan sumber data dapat dilakukan untuk mendapatkan lebih banyak data opini pelanggan. Dengan jumlah opini yang lebih banyak maka pengetahuan yang didapatkan dari data tersebut akan lebih banyak untuk proses Analisis Sentimen opini pelanggan pariwisata pantai Malang Selatan. Sehingga penerapan kerangka kerja Analisis Sentimen bisa menghasilkan analisis yang akurat terkait sentimen opini pelanggan yang diberikan wisatawan terhadap pariwisata pantai Malang Selatan.
2. Penggunaan metode untuk memperbaiki penggunaan kata yang tidak baku dan translasi frasa bahasa asing dalam tahap *Text Preprocessing*. Hal ini bertujuan untuk mengoptimalkan jumlah *feature* yang akan digunakan untuk tahap klasifikasi menggunakan algoritme klasifikasi.
3. Pemilihan aspek yang akan dianalisis dalam proses Analisis Sentimen dapat menggunakan pendekatan ekstraksi aspek dengan metode *Part-Of-Speech Tagger* dengan tujuan mendapatkan aspek yang memang benar-benar ada dalam *dataset*.

## DAFTAR PUSTAKA

- Alpaydin, E., 2010. *Introduction to machine learning*, Cambridge, MA.
- Bangor, A., Kortum, P.T. & Miller, J.T., 2008. An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24(6), hh.574–594.
- Bangor, A., Kortum, P.T. & Miller, J.T., 2009. Determining what individual SUS scores mean: adding an adjective rating scale. *Journal of Usability Studies*, 4(3), hh. 144-123.
- Boeing, G. & Waddell, P., 2016. New Insights into Rental Housing Markets across the United States: Web Scraping and Analyzing Craigslist Rental Listings. *Journal of Planning Education and Research*, 37(4), hh.457–476
- Bucur, C., 2015. Using Opinion Mining Techniques in Tourism. *Procedia Economics and Finance*, 23, hh.1666–1673.
- Castrillo-Fernández, O., Web Scraping: Applications and Tools, European Data Portal. Tersedia di: [https://www.europeandataportal.eu/sites/default/files/2015\\_web\\_Scraping\\_applications\\_and\\_tools.pdf](https://www.europeandataportal.eu/sites/default/files/2015_web_Scraping_applications_and_tools.pdf) [Diakses 28 Mei 2018].
- Chen, C.-houh, Härdle Wolfgang & Unwin, A., 2007. *Handbook of data visualization*, Berlin: Springer.
- Dietterich, T.G., 2000. Ensemble Methods in Machine Learning. Multiple Classifier Systems Lecture Notes in Computer Science, pp.1–15.
- Dinas Pariwisata dan Kabupaten Malang. 2018. <http://disparbud.malangkab.go.id/> [Diakses 28 Mei 2018]
- Eckerson, W. 2005. *Deploying Dashboard and Scorecards*. TD WI Best Practices Report
- Ekawati. D., & Khodra, M.L., 2017, Aspect-based sentiment analysis for Indonesian restaurant reviews, *2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA)*, Denpasar, pp. 1-6.
- Fachrina., Zulfa and Widyanoro., D.H., "Aspect-sentiment classification in *Opinion Mining* using the combination of rule-based and machine learning," *2017 International Conference on Data and Software Engineering (ICoDSE)*, Palembang, 2017, pp. 1-6.
- Feldman, R. & Sanger, J., 2007. *The Text Mining Handbook: advanced approaches in analyzing unstructured data*, Cambridge: University Press.
- Few, S., 2006. *Information Dashboard design*, Beijing: OReilly.
- Han, J., Kamber, M. & Pei, J., 2012. *Data mining: concepts and techniques*, Amsterdam: Elsevier/Morgan Kaufmann.



- Hercig, T. et al., 2016. UWB at SemEval-2016 Task 5: Aspect Based Sentiment Analysis. Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). Workshop on Semantic Evaluation, San Diego.
- Kementerian Pariwisata, 2014. Rencana Strategis Kementerian Pariwisata. *Kementerian Pariwisata*. Tersedia di: <http://www.kemenpar.go.id/asp/ringkasan.asp?c=14> [Diakses 19 Februari 2018].
- Liang, Y.D., 2013. *Introduction to programming using Python*, Boston: Pearson.
- Librian, A., 2017. sastrawi. *GitHub*. Tersedia di: <https://github.com/sastrawi/sastrawi> [Diakses 28 Mei 2018].
- Liu, B., 2012. *Sentiment analysis and Opinion Mining*, S.I.: Morgan & Claypool.
- Liu, B., 2013. *Web Data Mining: exploring hyperlinks, contents, and usage data*, Berlin: Springer.
- Mabroukeh, N.R. & Ezeife, C.I., 2010. A taxonomy of sequential pattern mining algorithms. *ACM Computing Surveys*, 43(1), pp.1–41.
- Malik, S. 2005. *Enterprise Dashboard – Design and Best Practices for IT*. John Wiley & Sons
- Manning, C.D., Raghavan, P. & Schütze Hinrich, 2009. *Introduction to information retrieval*, New York: Cambridge University Press.
- Marrese-Taylor, E. et al., 2013. Identifying Customer Preferences about Tourism Products Using an Aspect-based Opinion Mining Approach. *Procedia Computer Science*, 22, pp.182–191.
- Mitchell, R., 2018. *Web scraping with Python: collecting data from the modern web*, Beijing: O'Reilly.
- Moore, H.F., 1988. *Hubungan masyarakat: prinsip, kasus dan masalah (satu & dua)*, Bandung: Remadja Karya CV.
- Nathania, D.Z., Indriati & Bachtiar, F.A., 2018. Klasifikasi Spam Pada Twitter Menggunakan Metode Improved K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(10), pp.3948–3956.
- NLTK Project, 2018. Natural Language Toolkit. *Natural Language Toolkit*. Tersedia di: <https://www.nltk.org/> [Diakses 28 Mei 2018].
- Pang, B., Lee, L. & Vaithyanathan, S., 2002. Thumbs up? *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - EMNLP 02*.
- PD Jasa Yasa, Rencana Strategis. *PD JASA YASA Pemkab Malang*. Tersedia di: <http://jasayasa.malangkab.go.id/konten-23.html> [Diakses 28 Mei 2018].
- Pedregosa et al., 2011. Scikit-learn: Machine Learning in Python, *JMLR* 12, hh. 2825-2830



- Plotly, 2018. Modern Visualization for the Data Era. *Graph Maker · Plotly Online*. Tersedia di: <https://plot.ly/> [Diakses 12 Juli 2018].
- Pozzi et al, 2016. Approval network: a novel approach for sentiment analysis in social networks. *World Wide Web*, 20(4), hh.831–854.
- Python Software Foundation, 2018. The Python Tutorial. *Python 3.6.5 documentation*. Tersedia di: <https://docs.python.org/3.6/tutorial/> [Diakses 28 Mei 2018].
- Rivers, W.L. & Mathews, C., 1988. *Ethics for the media*, Englewood Cliffs, NJ: Prentice Hall.
- Sauro, Jeff, 2011. “*Measuring Usability With The System Usability Scale (SUS)*”, <http://www.measuringu.com/sus.php>. [Diakses 12 Juli 2018]
- Sharma, A., Sharma, M.K. & Dwivedi, R.K., 2017. Literature Review and Challenges of Data Mining Techniques for Social Network Analysis. *Advances in Computational Sciences and Technology*, 10(5), pp.1377–1354.
- Socher, R. et al., 2013. Recursive deep models for semantic compositionality over a sentiment treebank in *Empirical Methods in Natural Language Processing*, Palo Alto
- Soenarjo, Djoenasih. 1997. *Opini Publik*. Yogyakarta : Liberty
- SysNucleus, 2017. WebHarvy Web Scraper. *Web Scraping Explained*. Tersedia di: <https://www.webharvy.com/articles/advantages-webharvy-web-scraper.html> [Diakses 28 Mei 2018].
- TripAdvisor Inc., Tentang TripAdvisor. *TripAdvisor*. Tersedia di: <https://tripadvisor.mediaroom.com/id-about-us> [Diakses 28 Mei 2018].
- Turney, P.2., 2001. Thumbs up or thumbs down? *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL 02*.
- United Nation World Tourism Organization, UNWTO Annual Report 2017, United Nation World Tourism Organization. Tersedia di: <http://publications.unwto.org/>.
- Utama, I.G.B.R., 2016. *Pengantar Industri Pariwisata Revisi.*, Yogyakarta: Deepublish.
- Wiebe, J.M., Bruce, R.F. & Ohara, T.P., 1999. Development and use of a gold-standard data set for subjectivity classifications. *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics -*.
- Wijayanti, R. & Arisal, A., 2017. Ensemble approach for sentiment polarity analysis in user-generated Indonesian text. *2017 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*.

Wright, A., 2009. Mining the Web for Feelings, Not Facts. *New York Times*. Tersedia di: <http://www.nytimes.com/2009/08/24/technology/internet/24emotion.html> [Diakses 19 Februari 2018].

